Learning Across Games*

Friederike Mengel†
University of Alicante

May 2007

Abstract

In this paper 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.

JOB MARKET PAPER

JEL Classification: C70, C72, C73.

Keywords: Game Theory, Bounded Rationality, Learning, Analogies.
Comments welcome !!

*This paper has benefitted enormously from discussions with my supervisor Fernando Vega Redondo. I also wish to thank Larry Blume, David Easley, Ani Guerdjikova, Christoph Kuzmics, Fabrice Le Lec as well as seminar participants in Alicante, Budapest (ESEM 2007), Cornell, Edinburgh, Hamburg (SMYE 2007), Leuven, Madrid, Stony Brook and Warwick (RES 2007) for helpful comments. Financial support from the Spanish Ministery of Education and Science (grant SEJ 2004-02172) is gratefully acknowledged.

† *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 each 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 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 exogenous 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. For most of the paper we use reinforcement learning as underlying model of how agents learn partitions and actions. At the end of the paper we consider other learning models and show that our results are robust.

To fix ideas think about 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 will 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 for arbitrarily small reasoning costs there is an equilibrium in which players see the two games as analogous and where the responder always receives a strictly positive share of the pie. 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]

[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.

- Nash equilibria in weakly dominated strategies that are unstable to learning in a single game can be stabilized by learning across games.
- 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.

We 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. Finally we show that our results are robust when alternative learning models are used. In particular we discuss several variants of stochastic fictitious play and evolution in population games.

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 show that our results are robust to changes in the underlying learning model. In Section 7 we discuss related literature. Section 8 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] 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 games $\gamma \in \Gamma$ share the same action set A_i. Players partition the set of all games into subsets of games they do not distinguish. 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 of Z, as partitions of higher cardinality are more costly. The parameter ξ gives an upper bound on reasoning costs. 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) $\forall\ Z, \xi > 0$: $0 < \Xi(Z, \xi) < \xi$.

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

(iii) $\xi < \left|\min_{A_1 \times A_2 \times \Gamma} \pi^i(a, \gamma)\right|$.

Reasoning costs are strictly increasing and unimportant relative to the smallest possible payoff $\min \pi^i(a, \gamma)$ from any of the games. The case of small costs is the most interesting one. With high reasoning costs new predictions arise trivially.[4]

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 a reinforcement (or sometimes called stimulus response) model based on Roth and Erev (1995).[5] In this 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 active attractions is given by MZ_l where Z_l is the cardinality of the partition the agent 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 $a^t = (a^{1t}, a^{2t})$. Payoffs are normalized to be strictly positive and finite.[6] After playing a game players will update their propensities and attractions taking into account the payoff obtained.

State

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

[4] In Section 6.2 we will come back to this assumption and discuss it somewhat more.

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

[6] 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

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

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 $\{f_j\}_{\gamma_j \in \Gamma}$ and classified into g_k^{it} according to G^{it}.

(iii) Players choose action a_m with probability

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

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 $\varepsilon_0 > 0$. In the analogy class just visited ε_0 is best interpreted as noise or experimentation.[7] As ε_0 has a bigger effect on smaller β, it increases the probability that "suboptimal" actions are chosen. In analogy classes not visited, it can be seen as reflecting forgetting. Note that if an analogy class is never visited, then, 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 leads to a reversal to the uniform distribution for action choices.[8]

(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.[9]

[7] 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) or Hopkins (2002).

[8] Of course noise could 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.

[9] 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 equation (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 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 $q^{it} = (q_l^{it})_{G_l \in \mathcal{G}}$ and let $x^t = (x^{1t}, x^{2t}) \in \mathbf{X}$. The main interest lies in the evolution of x^t. $\mathbf{X}$ is the space in which these probabilities evolve. It has dimension $(2^J - 1)(M_1 + M_2 - 2) + 2(L - 1)$, where $M_1 = \operatorname{card} A_1$ and $M_2 = \operatorname{card} A_2$.[10]

3.1 Mean Dynamics

Action Choice and Phenotypic Play

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

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

- *Phenotypic play* in any game γ_j is described by the probabilities $\sigma_j^{it} = (\sigma_{1j}^{it}, ..., \sigma_{Mj}^{it})$ defined over the set of games Γ. The σ_j^{it} do not characterize an agents choice but how an agent actually behaves in a given game.

In other words the phenotypic 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: $\sigma_{mj}^{it} := \sum_{G_l \in \mathcal{G}} q_l^{it} \sum_{g_k \in G_l} p_{mk}^{it} \mathrm{I}_{jk}$ where $\mathrm{I}_{jk} = 1$ if $\gamma_j \in g_k$ and zero otherwise.

Mean Motion

[10] 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 all probabilities have to sum to one.

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.[11] 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 visited. $\sum_{G_l \in \mathcal{G}} q_l^{it} \mathrm{I}_{kl}$ 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 $S_l^{it}(x^t) = \Pi_l^{it}(x^t) - \sum_{G_h \in \mathcal{G}} q_h^{it} \Pi_h^{it}(x^t)$ where $\Pi_l^{it}(x^t)$ is the expected payoff net of reasoning costs obtained when using partition G_l.

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

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

Proof. Appendix A. ■

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

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.[12] 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 $\sum_{t=1}^{\infty}\left(\frac{1}{\beta_k^{it}}\right)^2 < \infty, \forall g_k \in \mathcal{P}^+(\Gamma), i = 1, 2$). 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.[13] One possibility to deal with this problem is to introduce additional parameters that take account of the relative speed of learning. We focus on a simpler way of dealing with this problem that consists in normalizing the process.[14]

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.[15] 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:*

$$\dot{p}_{mk}^{i} = p_{mk}^i r_k^i S_{mk}^i(x) + \varepsilon_0(1 - M p_{mk}^i) \tag{8}$$

$$\dot{q}_{l}^{i} = q_l^i S_l^i(x) + \varepsilon_1(1 - L q_l^i). \tag{9}$$

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

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

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

[14] See Hopkins (2002) or Laslier, Topol and Walliser (2001) 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.

[15] 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 (8)-(9).[16] More precisely let us denote the vector field associated with the system (8)-(9) 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 $F(x(t))$.[17] 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 restpoint of $F(x(t))$, then $\Pr\{\lim_{t\to\infty} x^t = x^*\} > 0$.[18] In the following analysis we will thus focus on the asymptotically stable points of (8)-(9).

4 Equilibrium Actions

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

(i) $\varepsilon_0 \to 0$ and (ii) $\varepsilon_1 = \lambda\varepsilon_0$ for some constant λ.

Noise is assumed to be vanishingly small and of the same order for both action and partition choices. The second condition ensures that there are no partitions whose choice probabilities converge faster to zero than noise ε_0. If this were the case noise would dominate in some analogy classes 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 restpoints $x^* = (p^*, q^*)$ 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 points of the system and the limit set $\lim_{\epsilon_0\to 0} E(\varepsilon_0) =: E^*$.

Proposition 2 *There exists $\overline{\xi}(\Gamma) > 0$ s.th. whenever $\xi < \overline{\xi}(\Gamma)$ any asymptotically stable point $x^* \in E^*$ must induce phenotypic 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.[19]

Naturally now the question arises how learning across games selects between (possibly) many Nash equilibria ? We will see in the following subsections that

[16] Equations (8)-(9) constitute a 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 studied unperturbed reinforcement learning.

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

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

[19] 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.

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. In these games weakly dominated strategies in the strategic form representation 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. Assume 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 again and so on. 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 she wants to keep) and b^κ the acceptance threshold. Let $a^{i\kappa}$ and $b^{i\kappa}$ be from the finite grid $A = \{0, \frac{1}{M}, \frac{2}{M}, ..\frac{M-1}{M}, 1\}$.[20] We focus on stationary strategies, i.e. strategies that do not depend on the decision node κ.

Assume that the players face two Rubinstein games that differ in the discount factor δ_j. At each point in time one game is randomly drawn and classified by the

[20] Assume that the grid A is fine enough s.t. it contains all equilibrium strategies described below.

agents into an analogy class according to the partition they hold. 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 game γ_1 has discount factor $\delta_1 \to 1$, and game γ_2 has discount factor $\delta_2 = 0$. Game γ_2 is essentially an Ultimatum Game, where the whole pie is gone of the first offer is not accepted. 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 herself and player 2 accepting all offers of at least $\frac{1}{M}$.[21] 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 (γ_1), the Ultimatum game (γ_2) and the coarse partition. Whenever there is no reasoning cost ($\Xi(Z, \xi) = 0, \forall Z \in \mathbb{N}$) all asymptotically stable restpoints 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) things change. Remembering that f_1 denotes the frequency with which game γ_1 occurs, we can state the following result.

Claim 1 *$\forall \xi > 0$ there exists an asymptotically stable point for $\Gamma_1 = \{\gamma_1, \gamma_2\}$ 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. 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

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

subjects often do not behave in accordance with subgame perfection.[22] 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 point $x^* \in E^*$.

(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. Generically in finite extensive form games with perfect information any equilibrium in weakly dominated strategies of the associated strategic form fails to be subgame perfect in the extensive form.[23] 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, while being sufficient, is not necessary to stabilize such an equilibrium.

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

Interesting predictions of learning across games can also arise if Γ contains games with mixed strategy equilibria. Again we start this subsection with an intuitive

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

[23] See chapter 6 in Osborne and Rubinstein (1994). Note that the qualifier generic here refers to the extensive form.

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_2 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.[24] If both players have matrix M' the resulting game γ_2 is an (Anti)-Coordination Game.[25] 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 players 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.[26] Simultaneous learning of actions and partitions leads to the same prediction whenever there are no reasoning costs ($\Xi(Z, \xi) = 0, \forall Z \in \mathbb{N}$). Even for vanishingly small costs things 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 point for Γ_2 involves both players holding the coarsest partition $G_C = \{\gamma_1, \gamma_2, \gamma_3\}$ with asymptotic probability 1 and choosing $p_c^* = 1/2$.*

Proof. Appendix B. ■

At the unique stable point, both players hold the coarse partition and play the mixed Nash equilibrium strategies. The intuition is as follows. 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

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

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

[26] This is shown in Appendix B. See also Ellison and Fudenberg (2000).

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

It is a well known result that strict Nash equilibria are asymptotically stable to any deterministic payoff monotone dynamics for a single game. Learning across games can sometimes destabilize strict Nash equilibria in Coordination and (Anti)-Coordination games 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 $\widehat{f}(\Gamma) > 0$ s.t. if $f_1/f_2 < \widehat{f}(\Gamma)$ the strict equilibrium $\widehat{\sigma}_1$ is not phenotypically induced at any asymptotically stable point $x^* \in E^*$.

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.[27] If there are no reasoning costs ($\Xi(Z, \xi) = 0, \forall Z \in \mathbb{N}$) any strict Nash equilibrium can be induced at an asymptotically stable point 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 mixed equilibrium in a different game, the strict equilibrium will be destabilized. The reason is that a) the mixed equilibrium will be observed in the second game at any asymptotically stable point 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 be tendency for agents to see the games as analogous and to

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

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

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.

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 point $x^* \in E^*$.*

(ii) If $\operatorname{card} \Gamma > 1$ this need not be true. Specifically let $\operatorname{card} \Gamma = 2$ and let γ_2 have an equilibrium $\widehat{\sigma}_2 = \widehat{\sigma}_1$ stable to learning in a single game. Then whenever $f_1/f_2 > \widehat{f}(\Gamma)$, 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.[28] 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

As we have noted before our perspective on partition learning is a very instrumental one. Rather than asking which games do agents *a priori perceive as analogous* (according to some exogenous similarity measure), we are interested in the question which games will agents *learn* to discriminate ? Consider for

[28] 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.

example the following three games occurring with the same frequency,

$$\gamma_1 : \begin{pmatrix} 3,5 & 3,4 \\ 1,4 & 4,3 \end{pmatrix}, \gamma_2 : \begin{pmatrix} 3,3 & 3,4 \\ 1,3 & 4,5 \end{pmatrix}, \gamma_3 : \begin{pmatrix} 1,\frac{1}{5} & 0,\frac{1}{7} \\ \frac{1}{4},\frac{1}{7} & \frac{1}{2},\frac{1}{5} \end{pmatrix}.$$

It is not clear what kind of a priori similarity criterion one should apply to the set of games $\Gamma = \{\gamma_1, \gamma_2, \gamma_3\}$. Games γ_1 and γ_2 are relatively closer in payoff space, but all three games are strategically different.[29] In fact the row player would like to match the opponent's action in all games but the column player has a dominant strategy to play the first (left) action in game γ_1 and the second (right) action in γ_2.[30] Now as on outcome of learning across games both players could either hold partition $\{\gamma_1, \{\gamma_2, \gamma_3\}\}$ or $\{\gamma_2, \{\gamma_1, \gamma_3\}\}$, but γ_1 and γ_2 will always be distinguished in equilibrium. The reason is that the supports of the sets of Nash equilibria in games γ_1 and γ_2 are disjoint.

In general whether any two games will be seen as analogous as an outcome of the learning process will depend on the degree of "overlap" between the Nash equilibria of the different games contained in Γ. Denote $S^{Nash}(\gamma_j)$ the support of the set of Nash equilibria $E^{Nash}(\gamma_j)$ of a game γ_j. Formally $S^{Nash}(\gamma_j) = \{a^i_m | \exists \sigma^i_j \in E^{Nash}(\gamma_j) \text{ with } \sigma^i_{mj} > 0\}$. 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 *There exists $\overline{\xi}(\Gamma) > 0$ s.th. whenever $\xi < \overline{\xi}(\Gamma)$ the finest partition G_F will be chosen with asymptotic probability $q^{i*}_F = 1, \forall i = 1, 2$ at all asymptotically stable points 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.*

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 phenotypic play for one of the players in one of the games. This player will gain from distinguishing these games. The following remark establishes an upper bound on the cardinality of the partitions agents will use in equilibrium.

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

Any partition of higher cardinality will either contain two different analogy classes in which the same pure action is chosen. Or - if a mixed action is chosen

[29] Rubinstein (1988) uses distance in payoff (or probability) space as a similarity criterium in one-person decision problems. Steiner and Stewart (2006) use such a criterium for games.

[30] Note also that if the row player held partition $\{\{\gamma_1, \gamma_2\}, \gamma_3\}$ (choosing the first action in $\{\gamma_1, \gamma_2\}$) the payoff information he would receive could only contradict such an analogy partitioning after sufficiently many trembles. A similar idea underlies the concept of subjective Nash equilibrium by Kalai and Lehrer (1995).

in some analogy class g_k - there will exist another analogy class $g_h \neq g_k$ in which a best response to the phenotypic play of the opponent is chosen $\forall \gamma_j \in g_k$. As merging these analogy classes will save reasoning costs, such a restpoint can never be stable.

6 Extensions

There are two features of our model that we would like to discuss somewhat more. The first point we would like to make is that the predictions for action choices that arise with learning across games are robust and continue to hold if other (and a priori quite different) learning models are considered. Secondly we would like to shortly discuss our assumption of small reasoning costs.

6.1 Other Learning Models

6.1.1 Stochastic Fictitious Play

The other model (apart from reinforcement learning) that has received a lot of attention in the literature is the model of stochastic fictitious play.[31] In stochastic fictitious play a group of players repeatedly play a normal form game. During each time period each player plays a best response to the time average of her opponent's play, but only after her payoffs have been randomly perturbed. In applying stochastic fictitious play to our context of *simultaneous* learning of actions and partitions two cases arise depending on whether or not players are able to correlate their action and partition choices. Before describing the choice rules in each of these cases let us introduce some notation. Denote

$$z_m^{it}(g_k^{-i}) = \frac{\sum_{\tau=1}^{t-1} \delta_m^i(\tau)\delta_k^{-i}(\tau)}{\sum_{\tau=1}^{t-1} \delta_k^{-i}(\tau)} \tag{10}$$

the frequency vector that describes the historical frequency of player i choosing action a_m whenever player $-i$ visits analogy class g_k. $\delta_m^i(\tau)$ takes the value 1 if player i chooses a_m at time τ and 0 otherwise. $z^{it}(g_k^{-i}) = (z_1^{it}(g_k^{-i}), ...z_{M_i}^{it}(g_k^{-i}))$ is the belief of player $-i$ about player i's action choice in the games contained in g_k. In the same spirit denote

$$\overline{\Pi}_l^{it}((x_\tau)_{\tau=1}^{t-1}, Z_l) = \frac{\sum_{\tau=1}^{t-1} \left[\pi^i(a^\tau, \gamma^\tau) - \Xi(Z_l)\right] \delta_l^i(\tau)}{\sum_{\tau=1}^{t-1} \delta_l^i(\tau)} \tag{11}$$

the historical (net) payoff obtained on average when visiting partition G_l.

Let us start with the case that seems closest to reinforcement learning, where agents do not have the possibility to correlate their action and partition choices. According to fictitious play the player first picks the partition with the highest expected payoff which (as players do not correlate action and partition choice)

[31] See for example Fudenberg and Levine (1998). Hopkins (2002) compares the long run behavior of reinforcement learning and stochastic fictitious play.

is equivalent to simply choosing q^{it*} to maximize $q^{it}\overline{\Pi}_l(\cdot)$ where $\overline{\Pi}_l(\cdot)$ describes player i's historical payoff given $i's$ and $-i$'s action choices. The choice rule for partitions is

$$q^{it*} \in \arg\max \sum_{G_h \in \mathcal{G}} q_h^t \overline{\Pi}_h(\cdot) + \varepsilon_1 \varphi(q^t) \tag{12}$$

where $\varphi(q)$ is a deterministic perturbation.[32] Some restrictions on the shape of this function are given in Appendix C. As the player's payoffs do not directly depend on their opponent's partition choice (only indirectly through the induced actions) and as they do not correlate action and partition choice the latter is entirely non-strategic. Given their partition choice agents then choose their actions for a given analogy class as follows,

$$p_k^{it*} \in \arg\max p_k^{it} \left[\sum_{\gamma_j \in g_k^i} f_j \pi(\gamma_j) \right] z^{-it}(g_k^i) + \varepsilon_0 \varphi(p_k^t). \tag{13}$$

where $\pi(\gamma_j)$ is the payoff-matrix of game γ_j. Agents use the average payoff matrix across all the games contained in analogy class g_k as relevant information.[33]

If agents are able to correlate partition and action choice we have the following choice rule,

$$(q^{it*}, p_k^{it*}) \in \arg\max \sum_{G_h \in \mathcal{G}} q^{it} \left[\sum_{g_k \in G_l} p_k^{it} \left[\sum_{\gamma_j \in g_k^i} f_j \pi(\gamma_j) \right] z^{-it}(g_k^i) \right] \mathrm{I}_{jk} + \varepsilon \varphi(p^t, q^t), \tag{14}$$

where I_{jk} takes the value 1 if $\gamma_j \in g_k$ and zero otherwise. Given that partition choice is not directly payoff relevant (only through the action choices it induces), it is not surprising that correlating both choices does not fundamentally change the results. In both cases (correlation and non-correlation) stochastic fictitious play gives rise to differential equations that coincide with those associated with the reinforcement learning process up to a multiplicative constant and a difference in the noise term. We can state the following proposition.[34]

Proposition 7 *Under stochastic fictitious play with choice rules (12) - (13) or (14) Propositions* 2-6 *as well as Claims* 1-2 *continue to hold.*

Proof. Appendix C. ■

Next we want to show that - while the results are robust to changes in the underlying model - the notion of analogy employed can be crucial.

[32] Hofbauer and Sandholm (2002) have shown that for any stochastic perturbation used in (12) there is always an alternative representation using a deterministic noise function.

[33] In the terminology of Germano (2007) the matrix across games $\sum_{\gamma_j \in g_k^i} f_j \pi(\gamma_j)$ would be the "average game".

[34] It is not new to the literature that stochastic fictitious play and reinforcement learning can lead to similar ODE's in the stochastic approximation. See Benaim and Hirsch (1999) or Hopkins (2002) among others.

6.1.2 Stochastic Fictitious Play with Analogy Based Expectations

Jehiel (2005) has proposed a (static) model where seeing two games as analogous only means having the same expectations about the opponent's behavior. This implies that action choice can still be different even in games that are seen as analogous. In this section we use Jehiel's (2005) notion of analogy and add an endogenous partition choice relying on the stochastic fictitious play algorithm.

Then in the case of no correlation choice rule (13) is replaced by,

$$p_j^{it*}(g_k) \in \arg\max p_j^{it}(g_k)\pi(\gamma_j)z^{-it}(g_k^i) + \varepsilon_0\varphi(p_{jk}). \tag{15}$$

Note that the choice variable here is $p_j^{it}(g_k)$ instead of p_k^{it} in equation (13). With analogy based expectations action choice is conditioned on both the game *and* the analogy class the game is contained in. Agents choose a best response to their beliefs $z^{-it}(g_k^i)$ (that depend on the analogy class) in each game separately.

If agents are able to correlate partition and action choice the choice rule is as follows,

$$(q^{it*}, p_j^{it*}(g_k)) \in \arg\max \sum_{G_h \in \mathcal{G}} q^{it} \sum_{g_k \in G_l} \sum_{\gamma_j \in \Gamma} f_j \left[p_j^{it}(g_k)\pi(\gamma_j)z^{-it}(g_k^i)\right] + \varepsilon\varphi(p^t, q^t). \tag{16}$$

These processes are quite different from what we have considered until now, as a different notion of analogy is used. And of course the ODE's associated with either of them will not coincide with (8) - (9). What we are interested in is whether the phenotypic play of the agents will be such that the results derived above continue to hold. The next proposition shows that - maybe not surprisingly - the predictions of such a model do not always coincide with the predictions of our model.

Proposition 8 *Under stochastic fictitious play with choice rules (12) - (15) or (16) Proposition 2 continues to hold. On the other hand there are conditions under which Propositions 3-5 fail.*

Proof. Appendix C. ■

Proposition 8 shows that - while the results are robust to changes in the underlying learning model - the notion of analogy employed can be crucial. With Jehiel's (2005) notion of analogy Propositions 3-5 continue to hold only if additional restrictions are met. The proposition also illustrates the discipline that endogenizing partition choice imposes on the process. The deviations from Nash equilibrium that Jehiel (2005) observes do not occur when partition choice is endogenous (and reasoning costs small). To illustrate this point consider the following example taken from Jehiel (2005).

Example I

Consider the following games occurring with the same frequency,

γ_1 :

	L	LM	RM	R
H	$5,2$	$0,2$	$2,4$	$0,0$
L	$4,3$	$3,0$	$1,0$	$2,0$

, γ_2 :

	L	LM	RM	R
H	$3,0$	$4,2$	$2,0$	$1,1$
L	$0,2$	$5,2$	$0,0$	$2,4$

.

The unique Nash equilibrium is (H, RM) in γ_1 and (L, R) in γ_2. Jehiel (2005) shows that the following is an analogy-based expectations equilibrium. Player 1 sees both games as analogous and plays L in game γ_1 and H in γ_2 best responding to beliefs $z^1(\{\gamma_1, \gamma_2\}) = (\frac{1}{2}, \frac{1}{2}, 0, 0)$. Player 2 distinguishes the games and plays L in γ_1 and LM in γ_2 best responding to beliefs $z^2(\{\gamma_1\}) = (0, 1)$ and $z^2(\{\gamma_2\}) = (1, 0)$. This action profile is not a Nash equilibrium in either game. Endogenizing partition choice with the stochastic fictitious play process though shows that such a point cannot be stable (for small reasoning costs). Consider the partition choice of player 1. In the off-equilibrium analogy classes $\{\gamma_1\}$ and $\{\gamma_2\}$ beliefs will eventually converge to $z^1(\{\gamma_1\}) = (1, 0, 0, 0)$ and $z^1(\{\gamma_2\}) = (0, 1, 0, 0)$. Whenever player 1 holds the fine partition she will choose H in γ_1 and L in γ_2 giving her a payoff of 5 in both games (as opposed to 4 with the coarse partition). Thus the historical (net) payoff obtained when visiting the fine partition G_F will converge to 5. Under either choice rule (12) - (15) or (16) player 1 will eventually start to use the fine partition, destabilizing such a restpoint. Note though that if player 1 were forward-looking (anticipating the final outcome) she might prefer using the coarse partition.[35]

6.1.3 Population Games

Hofbauer and Sandholm (2007) have introduced a model of evolution in population games with randomly perturbed payoffs.[36] Just like stochastic fictitious play the expected motion in their model is described by the perturbed best response dynamics. The main difference between their model and the model of stochastic fictitious play lies in the definition of the state variable. As Hofbauer and Sandholm study evolution in population games the state variable in their model is the proportion of players that choose a certain strategy. To establish convergence they thus cannot just consider the time average of play, but instead have to take first the limit as the population size grows to infinity. This has as a consequence that the their model selects in general more strongly than stochastic fictitious play. While the two models do not always lead to the same predictions, they do so often. In particular whenever stochastic fictitious play converges to a unique restpoint, their dynamics also does (as it happens e.g. in Claim 2). While analyzing learning across games in population games is beyond the scope of this paper their results give us confidence that many of our results will extend.

[35] This suggests that any learning foundation for analogy-based expectation equilibrium with endogenous partitions should involve some degree of forward looking behavior.

[36] See also Benaim and Weibull (2003), Blume (1993) or Young (1998).

6.2 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. If this were the case one could argue on evolutionary grounds that reasoning costs will most likely tend to be small. The following simple example shows that having smaller reasoning costs need not always lead to better outcomes for a player.

Example II

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 point 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 she would see both games as analogous.[37] 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 she has high reasoning costs.

This example shows that 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.[38]

7 Related Literature

The idea that similarities or analogies play an important role for economic decision making has long been present in the literature.[39] 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

[37] Of course we have defined the process only for small reasoning costs (relative to the game payoffs). Extending to general costs is no problem though. See footnote 9.

[38] There is some literature related to this issue. See for example Robson (2001) and the references contained therein.

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

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.[40] 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 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 our 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.[41]

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.[42]

8 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

[40] 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.

[41] Other papers in the automaton tradition investigating equilibria in the presence of complexity costs are Abreu and Rubinstein (1988) or Eliaz (2003). Germano (2007) studies the evolution of rules for playing stochastically changing games.

[42] Pemantle (1990), Benaim and Hirsch (1999) or Benaim and Weibull (2003) are also technically related.

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. One possible explanation for this phenomenon could be that different frames trigger different analogies. We conjecture 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] Blume, L. (1993), The statistical mechanics of strategic interaction, Games and Economic Behavior, 5, 387-424.

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

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

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

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

[11] 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.

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

[13] Germano, F. (2007), Stochastic Evolution of Rules for Playing Finite Normal Form Games, Theory and Decision 62 (4), 311-333.

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

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

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

[17] Hofbauer, J. and W. Sandholm (2002), On the global convergence of stochastic fictitious play, Econometrica 70, 2265-2294.

[18] Hofbauer, J. and W. Sandholm (2007), Evolution in games with randomly perturbed payoffs, Journal of Economic Theory 132, 47-69.

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

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

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

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

[23] Kalai, E. and E. Lehrer (1995), Subjective Games and Equilibria, Games and Economic Behavior 8, 123-163.

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

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

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

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

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

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

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

[31] Robson, A. (2001), Why would Nature give Individuals Utility Functions ?, Journal of Political Economy 109 (41), 900-914.

[32] 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.

[33] 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.

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

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

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

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

[38] Young, P. (1998), Individual Strategy and Social Structure, Princeton: Princeton University 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,

$$p_{mk}^{1(t+1)} - p_{mk}^{1t} = \begin{cases} \frac{\beta_{mk}^{1t}+\pi^1(a^t,\gamma^t)+\varepsilon_0}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+\pi^1(a^t,\gamma^t)+M\varepsilon_0} - \frac{\beta_{mk}^{1t}}{\sum_{a_h\in A_1}\beta_{hk}^{1t}} & \text{if } g_k, a_m \in w^{it} \\ \frac{\beta_{mk}^{1t}+\varepsilon_0}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+\pi^1(a^t,\gamma^t)+M\varepsilon_0} - \frac{\beta_{mk}^{1t}}{\sum_{a_h\in A_1}\beta_{hk}^{1t}} & \text{if } g_k \in w^{it} \\ & \quad a_m \notin w^{it} \\ \frac{\beta_{mk}^{1t}+\varepsilon_0}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+M\varepsilon_0} - \frac{\beta_{mk}^{1t}}{\sum_{a_h\in A_1}\beta_{hk}^{1t}} & \text{if } g_k \notin w^{it} \end{cases} \quad (17)$$

or equivalently

$$p_{mk}^{1(t+1)} - p_{mk}^{1t} = \begin{cases} \frac{(1-p_{mk}^{1t})\pi^1(a^t,\gamma^t)+\varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+\pi^1(a^t,\gamma^t)+M\varepsilon_0} & \text{if } g_k, a_m \in w^{it} \\ \frac{-p_{mk}^{1t}\pi^1(a^t,\gamma^t)+\varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+\pi^1(a^t,\gamma^t)+M\varepsilon_0} & \text{if } g_k \in w^{it} \\ & a_m \notin w^{it} \\ \frac{\varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+M\varepsilon_0} & \text{if } g_k \notin w^{it} \end{cases} . \quad (18)$$

The first event has the following probability $\sum_{\gamma_j\in\Gamma} f_j \, \mathrm{I}_{jk} \sum_{G_l\in\mathcal{G}} q_l^{1t}\mathrm{I}_{kl} p_{mk}^{1t} \sum_{a_n\in A_2}\left(\sum_{G_l\in\mathcal{G}} q_l^{2t} \sum_{g_k\in G_l} p_{nk}^{2t}\mathrm{I}_{jk}\right)$ where $\mathrm{I}_{jk}(\mathrm{I}_{kl})$ = 1 if $\gamma_j \in g_k$ ($g_k \in G_l$) and zero otherwise.[43] The second event has probability $\sum_{\gamma_j\in\Gamma} f_j \; \mathrm{I}_{jk} \sum_{G_l\in\mathcal{G}} q_l^{1t}\mathrm{I}_{kl} \sum_{a_h\in A} p_{hk}^{1t}(1-\delta_{hm}) \sum_{a_n\in A_2}\sigma_{nj}^{2t}$ where δ_{hm} is the Kronecker delta.[44]

The third event has probability $\sum_{\gamma_j\in\Gamma} f_j(1-\mathrm{I}_{jk}) + f_j\mathrm{I}_{jk} \sum_{G_l\in\mathcal{G}} q_l^{1t}(1-\mathrm{I}_{kl})$. Summing over all possible events (weighted with the probabilities) gives the mean change:

$$\begin{aligned} \left\langle p_{mk}^{1(t+1)} - p_{mk}^{1t} \right\rangle &= \sum_{\gamma_j\in g_k} f_j \sum_{G_l\in\mathcal{G}} q_l^{1t}\mathrm{I}_{kl} \\ &\left[p_{mk}^{1t} \sum_{a_n\in A_2} \frac{(1-p_{mk}^{1t})\pi^1(a_m^1,a_n^2,\gamma_j)\sigma_{nj}^{2t}+\varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+\pi^1(a_m^1,a_n^2,\gamma_j)+M\varepsilon_0} \right. \\ &\left. + \sum_{a_\eta\neq a_m\in A_1} p_{\eta k}^{1t} \sum_{a_n\in A_2} \frac{-p_{mk}^{1t}\pi^1(a_\eta^1,a_n^2,\gamma_j)\sigma_{nj}^{2t}+\varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+\pi^1(a_h^1,a_n^2,\gamma_j)+M\varepsilon_0} \right] \\ &+ \left(1 - \sum_{\gamma_j\in g_k} f_j \sum_{G_l\in\mathcal{G}} q_l^{1t}\mathrm{I}_{kl}\right) \frac{\varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+M\varepsilon_0} \end{aligned} \quad (19)$$

Denoting $\beta_k^{1t} = \sum_{a_h\in A_1}\beta_{hk}^{1t}$ this can be rewritten concisely as follows,

$$\left\langle p_{mk}^{1(t+1)} - p_{mk}^{1t} \right\rangle = \frac{1}{\beta_k^{1t}}[p_{mk}^{1t} r_k^{1t} S_{mk}^{1t}(\cdot) + \varepsilon_0(1-Mp_{mk}^{1t})] + O\left((\frac{1}{\beta_k^{1t}})^2\right). \quad (20)$$

To see that the difference between the first term in (20) and expression (19) is indeed of order $(\frac{1}{\beta_k^{1t}})^2$ note that,

$$\frac{p_{mk}^{1t} r_k^{1t} S_{mk}^{1t}(\cdot) + \varepsilon_0(1-Mp_{mk}^{1t})}{\beta_k^{1t}} - \left\langle p_{mk}^{1(t+1)} - p_{mk}^{1t} \right\rangle$$

[43] Note that $\left(\sum_{G_l\in\mathcal{G}} q_l^{2t} \sum_{g_k\in G_l} p_{nk}^{2t}\mathrm{I}_{jk}\right) = \sigma_{nj}^{2t}$.

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

$$= \frac{p_{mk}^{1t} r_k^{1t} S_{mk}^{1t} + \varepsilon_0(1 - M p_{mk}^{1t}) - p_{mk}^{1t} r_k^{1t} S_{mk}^{1t}(1 + \frac{\pi^1(\cdot) + M\varepsilon_0}{\beta_k^{1t}})^{-1}}{\beta_k^{1t}}$$
$$- \frac{\varepsilon_0(1 - M p_{mk}^{1t})(1 + \frac{M\varepsilon_0}{\beta_k^{1t}})^{-1}}{\beta_k^{1t}}$$

$$= p_{mk}^{1t} r_k^{1t} S_{mk}^{1t} \frac{(\pi^1(\cdot) + M\varepsilon_0)}{\beta_k^{1t}(\beta_k^{1t} + \pi^1(\cdot) + M\varepsilon_0)} + \varepsilon_0(1 - M p_{mk}^{1t}) \frac{M\varepsilon_0}{\beta_k^{1t}(\beta_k^{1t} + M\varepsilon_0)}.$$

■

Proof of Lemma 2:

Proof. The changes in partition choice probabilities are given by

$$q_l^{1(t+1)} - q_l^{1t} = \begin{cases} \frac{(1 - q_l^{1t})\left(\pi^1(a^t, \gamma^t) - \Xi(Z_l)\right) + \varepsilon_1(1 - L q_l^{1t})}{\sum_{G_h \in \mathcal{G}} \alpha_h^t + \pi^1(a^t, \gamma^t) + L\varepsilon_1} & \text{if } G_l \in w_1^t \\ \frac{-q_l^{1t}\left(\pi^1(a^t, \gamma^t) - \Xi(Z_h)\right) + \varepsilon_1(1 - L q_l^{1t})}{\sum_{G_h \in \mathcal{G}} \alpha_h^t + \pi^1(a^t, \gamma^t) + L\varepsilon_1} & \text{if } G_l \notin w_1^t \end{cases} \tag{21}$$

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,

$$\left\langle q_l^{1(t+1)} - q_l^{1t} \right\rangle = \frac{1}{\alpha^{1t}} [q_l^{it} S_l^{it}(x) + \varepsilon_1(1 - L q_l^{it})] + O\left((\frac{1}{\alpha^t})^2\right). \tag{22}$$

■

Proof of Proposition 1:

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

$$\begin{aligned} p_{mk}^{i(t+1)} &= p_{mk}^{it} + \frac{1}{\beta_k^{it}} \widetilde{Y}_{mk}^{it} \\ q_l^{i(t+1)} &= q_l^{it} + \frac{1}{\alpha^{it}} Y_l^{it} \end{aligned} \tag{23}$$

$\forall i = 1, 2, \forall a_m \in A_i, \forall g_k \in \mathcal{P}^+(\Gamma), \forall G_l \in \mathcal{G}$. The Y^{it} and $\widetilde{Y}^{it}$ can be decomposed as follows, $\widetilde{Y}_{mk}^{it} = \widetilde{y}_{mk}^i(x^t) + \widetilde{\omega}^{it}(c^t, d^t) + \widetilde{\upsilon}^{it}$ and $Y_l^{it} = y_l^i(x^t) + \omega^{it}(c^t, d^t) + \upsilon^{it}$. 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 $\widetilde{y}_{mk}^i(x^t) = p_{mk}^{it} r_k^{it} S_{mk}^{it}(\cdot) + \varepsilon_0(1 - M p_{mk}^{it})$ 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 unique step size of order t^{-1}, $\beta_k^{it} = \alpha^{it} = 1/(\mu + t\theta)$ can be substituted in (23). It can be verified that the following conditions hold for the normalized process. **(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 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 normalized process can be approximated by the deterministic system $\left(\dot{p}_{mk}^{\,i} = \widetilde{y}_{mk}^i(x)\right), \forall a_m \in A_i, g_k \in \mathcal{P}^+(\Gamma)$ and $\left(\dot{q}_l^{\,i} = y_l^i(x)\right), \forall G_l \in \mathcal{G}, i = 1, 2$ as standard results in stochastic approximation theory show.[45] ■

B Appendix: Proofs from Sections 4 and 5

Proof of Proposition 2:

Proof. We will show that no point $\widehat{x}$ that induces phenotypic play $(\sigma_j^1, \sigma_j^2) \notin E^{Nash}(\gamma_j)$ can be stable. As (σ_j^1, σ_j^2) is not a Nash equilibrium one player i will have a strictly better response a_m in some game γ_j. If γ_j is an element of a singleton analogy class g_k the claim is straightforward, as the expected payoff of a_m at $\widehat{x}$ conditional on visiting g_k is strictly higher than that of all actions on average, i.e. $S_{mk}(\widehat{x}) > 0$. It follows then directly from (8) that the growth rate function $\left(\dot{p}_{mk}/p_{mk}\right)$ of a_m is strictly positive for all x in an open neighborhood of $\widehat{x}$ (note that $\widehat{x}$ is interior because of the perturbation).

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

Consider next the case where γ_j is an element of a non-singleton analogy class. Denote $\phi := \pi^i(a_m, \sigma_j^{-i}, \gamma_j) - \pi^i(\sigma_j^i, \sigma_j^{-i}, \gamma_j) > 0$ the payoff loss incurred by choosing σ_j^i instead of the better response 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 $\exists \overline{\xi} < \phi$ such that $\forall \xi < \overline{\xi} : \Pi_l^i(\widehat{x}) - \Pi_h^i(\widehat{x}) = \phi - (\Xi(Z_l) - \Xi(Z_h)) > 0$ and thus the growth rate function $\dot{q}_l/q_l$ is strictly positive for all x in an open neighborhood of $\widehat{x}$. Consequently $\widehat{x}$ cannot be a stable restpoint. ■

Proof of Claim 1:

Proof. Consider the Rubinstein Bargaining game with discount factor $\delta = f_1$. This is the expected discount factor when both games are seen as analogous (and game γ_1 occurs with frequency f_1). Call this game the "average game". Note that player 1 chooses $a^1 = \frac{1}{1+f_1}$ and player 2 chooses $b^2 = \frac{f_1}{1+f_1}$ in any Nash equilibrium of the average game in which no player uses a strategy that is weakly dominated by some other pure strategy.[46] As (because of the perturbation) all restpoints are interior this implies that, given any payoff linear selection dynamics, the strategies $\left(a^1 = \frac{1}{1+f_1}, b^2 = \frac{f_1}{1+f_1}\right)$ will be observed with asymptotic probability one in the average game. (See for example Proposition 5.8 in Weibull (1995)).

Now we will show that there exists an asymptotically stable point x^* of the dynamics (8)-(9) in which both players hold the coarse partition (i.e. play the "average game") with asymptotic probability one and choose $a^{1*} = \frac{1}{1+f_1}$ and $b^{2*} = \frac{f_1}{1+f_1}$ when visiting analogy class $\{\gamma_1, \gamma_2\}$. First note that when visiting the "off equilibrium" analogy classes g_U and g_R the best response of player 1 is always to play $a^1 = a^{1*}$. The best response for player 2 is to play $b^2 = b^{2*}$ when visiting g_R, but she will end up randomizing between strategies (a, b) with $b \leq b^{2*}$ in g_U (as noise tends to drive the dynamics to the interior). While there can be (gross) gains of $O(\varepsilon_0)$ for player 2 if she uses another analogy partitioning (as $x^* \in \operatorname{int} \mathbf{X}$ and as b^{2*} is weakly dominated in the ultimatum game), there can be no net gains. More precisely let $\mathcal{N}_{x^*}$ be an open neighborhood of x^* and denote $\Xi(x)$ the total reasoning cost at x, i.e. $\Xi(x) = \sum_{G_l \in \mathcal{G}} q_l \Xi(Z_l)$. Then $\forall \xi > 0$,

$$\begin{aligned} & \sum_{\Gamma} \left(\pi^i(x^*, x^{-i}, \gamma_j) - \pi^i(x, \gamma_j)\right) - (\Xi(1) - \Xi(x)) \qquad (24) \\ = \quad & O(\varepsilon_0) - (\Xi(1) - \Xi(x)) > 0, \end{aligned}$$

[46] Note that in the Ultimatum game all strategies of player 2 are weakly dominated except the strategies $(a, 0)$ and $(a, \frac{1}{M})$. 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.

$\forall x \in \mathcal{N}_{x^*} \cap \operatorname{int} \mathbf{X}, i = 1, 2$. 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 (24) 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 $\operatorname{card} \Gamma = 1$ there is only one partition and one analogy class. Denote a_w^i the strategy that is weakly dominated by another strategy a_d^i for player i in game γ_1. Clearly $\pi^i(a^{*i}, x^{-i}, \gamma_1) - \pi^i(a_w^i, x^{-i}, \gamma_1) > 0, \forall x^{-i} \in \operatorname{int} \mathbf{X}_{-i}$. Consider a restpoint $\widehat{x}$ that induces a^{wi}. As $\widehat{x}$ is interior there exists a neighborhood $\mathcal{N}_{\widehat{x}}$ of $\widehat{x}$ s.t. $\pi^i(a^{*i}, x^{-i}, \gamma_1) - \pi^i(x, \gamma_1) + O(\varepsilon_0) > 0, \forall x \in \mathcal{N}_{\widehat{x}} \cap \operatorname{int} \mathbf{X}$ and consequently $\widehat{x}$ cannot be a stable restpoint.[47]

(ii) We will show that the restpoint x^* where both players hold the coarse partition $G_C = \{\gamma_1, \gamma_2\}$ with asymptotic probability $q_C^* = 1$ is asymptotically stable. Consider first action choice in $g_C = \{\gamma_1, \gamma_2\}$. For all x in an open neighborhood of x^* we have that $\sum_\Gamma \left(\pi^i(a^i, x^{-i}, \gamma_j) - \pi^i(x^*, \gamma_j)\right) + O(\varepsilon_0) < 0$, $\forall a^i \neq a_w^i$ and that $\sum_\Gamma \left(\pi^i(a_w^i, x^{-i}, \gamma_j) - \pi^i(x^*, \gamma_j)\right) + O(\varepsilon_0) > 0$, as a_w^i is a strict best response to x^{-i} in game γ_2 and a best response in γ_1. Next note that in all "off-equilibrium" analogy classes action choice will converge to a best responses to a_w^{-i} and consequently deviations in partition choice frequencies will lead at best to gains of order ε_0. But then there exists a neighborhood $\mathcal{N}'_{x^*}$ of x^* such that $\forall \xi > 0$, $\sum_\Gamma \left(\pi^i(\widehat{x}, x^{-i}, \gamma_j) - \pi^i(x, \gamma_j)\right) - (\Xi(1) - \Xi(x)) > 0$, $\forall x \in \mathcal{N}'_{x^*} \cap \mathbf{X}$, $i = 1, 2$. A strict Lyapunov function can be found as above. ■

Proof of Claim 2:

Proof. Let $G_1 = \{\{\gamma_1\}, \{\gamma_2\}, \{\gamma_3\}\}$, $G_2 = \{\gamma_1, \{\gamma_2, \gamma_3\}\}, G_3 = \{\gamma_2, \{\gamma_1, \gamma_3\}\}$, $G_4 = \{\gamma_3, \{\gamma_1, \gamma_2\}\}$ and $G_5 = \{\gamma_1, \gamma_2, \gamma_3\}$ be the five possible partitions of Γ_2. 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) First note that in analogy class $\{\gamma_3\}$ the unique Nash equilibrium strategy $\sigma_3 = 1/2$ will be observed at any asymptotically stable point. Also note that any action is a best response to $\sigma^{-i} = 1/2$ in any of the games $\gamma \in \Gamma_2$. Now consider the Jacobian matrix $\mathcal{M}$ associated with the linearization of the dynamics at restpoints $\widehat{x}$ that involve $q_1^i > 0$, for some $i = 1, 2$. 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 the "off equilibrium" analogy class $\{\gamma_1, \gamma_3\}$. Consequently the diagonal element of $\mathcal{M}$, $\left(\partial \dot{q}_3^i / \partial q_3^i\right) = \Xi(\widehat{x}) - \Xi(2) - 5\varepsilon_1 > 0$, as the coarse partition must have probability zero at $\widehat{x}$ (and thus $\Xi(\widehat{x}) > \Xi(2)$). It can be seen analogously that $\left(\partial \dot{q}_2^i / \partial q_2^i\right) > 0$ if $f_2 > f_3$ and if $f_3 > \max\{f_1, f_2\}$ either $\left(\partial \dot{q}_3^i / \partial q_3^i\right) > 0$ or

[47] Part (i) of this proposition also follows from Proposition 5.8 in Weibull (1995) and the fact that (because of the perturbation) all restpoints are interior.

$\left(\partial \dot{q}_2^i / \partial q_2^i\right) > 0$. Instability of restpoints involving $q_4^i > 0$ for some $i = 1, 2$ is shown analogously. Neither can a stable restpoint involve $q_l^i > 0$ for $l = 2, 3$. If $f_3 > \min\{f_1, f_2\}$, player 1 will play a fully mixed strategy $p_4^1 = 1/2$ in $\{\gamma_2, \gamma_3\}$ and player 2 will play the mixed strategy $p_5^2 = 1/2$ in analogy class $\{\gamma_1, \gamma_3\}$. It then follows immediately by arguments analogous to those above that $G_2 \notin \text{supp } q^{1*}$ and $G_3 \notin \text{supp } q^{2*}$. Furthermore note that no restpoint where player 2 holds partition G_2 and player 1 partition G_3 can induce Nash play in all games and thus (by Proposition 2) can't be asymptotically stable. If $f_3 < \min\{f_1, f_2\}$ analogous arguments apply.

(ii) Now we will show that the restpoint where both players choose the coarsest partition and play the mixed strategy $p = 1/2$ is asymptotically stable. The payoff matrix of the "average" game is given by,

$$\begin{pmatrix} 2(f_1 + f_3) + f_2 & 2f_2 + f_1 + f_3 \\ 2f_2 + f_1 + f_3 & 2(f_1 + f_3) + f_2 \end{pmatrix} \text{ for player 1} \tag{25}$$

and

$$\begin{pmatrix} 2f_1 + f_2 + f_3 & 2(f_2 + f_3) + f_1 \\ 2(f_2 + f_3) + f_1 & 2f_1 + f_2 + f_3 \end{pmatrix} \text{ for player 2.} \tag{26}$$

Given the assumption that $f_j < 1/2$ for $j = 1, 2$ - (25) and (26) 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 (25) - (26). 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_j the term under the square root is negative and thus both eigenvalues have strictly negative real parts.[48] 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 $\text{sign}\left[O(\varepsilon_0)\right] \gtreqless 0 \Leftrightarrow p_{mk}^i \lesseqgtr \frac{1}{2}$ 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. (i) As $\text{card}\,\Gamma = 1$ there is trivially only one partition and one analogy class $g = \gamma_1$. But then part (i) of this proposition is a standard result. See for example Proposition 5.11 in

[48] Under the unperturbed dynamics all eigenvalues are purely imaginary in this class of games. Posch (1997) has shown that unperturbed reinforcement learning leads to cycling.

Weibull (1995).

(ii) Let the games have payoff matrices given by

$$\gamma_1: \begin{array}{|c|c|c|}\hline & H & L \\ \hline H & a_1, a_1 & a_2, a_3 \\ \hline L & a_3, a_2 & a_4, a_4 \\ \hline \end{array}, \gamma_2: \begin{array}{|c|c|c|}\hline & H & L \\ \hline H & b_1, c_1 & b_2, c_3 \\ \hline L & b_3, c_2 & b_4, c_4 \\ \hline \end{array}, \tag{27}$$

where $a_1 > a_3$ and $a_4 > a_2$. As we want γ_2 to have a mixed equilibrium that is stable to learning in a single game we assume without loss of generality that $b_1 > b_3, b_4 > b_2, c_1 < c_3$ and $c_4 < c_2$. (See part (ii) of the proof of Claim 2). The payoff matrix across games is given by

$$\begin{pmatrix} f_1a_1 + f_2b_1, f_1a_1 + f_2c_1 & f_1a_2 + f_2b_2, f_1a_3 + f_2c_3 \\ f_1a_3 + f_2b_3, f_1a_2 + f_2c_2 & f_1a_4 + f_2b_4, f_1a_4 + f_2c_4 \end{pmatrix}. \tag{28}$$

Whenever $f_1/f_2 < (c_3 - c_1)/(a_1 - a_3) =: \widehat{f}(\Gamma)$ 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 probability $q^{*i}_C > 0$, then we need to have $p^i_{HC} = p^i_{H1} = 1$. (The condition for phenotypic play in game γ_1 is $\sigma^i_{H1} = q^i_C\ p^i_{HC} + (1 - q^i_C)p^i_{H1} = 1$). In order to induce a Nash equilibrium also in game γ_2 one needs $p^i_{H2} = 0$ and $q^i_C = \sigma^i_{H2}$, where σ^i_{H2} is the equilibrium strategy in analogy class $g_2 = \{\gamma_2\}$. But then for any player either $p^i_{HC} = 1$ is not a best response to the phenotypic play of player $-i$ or the diagonal element $\left(\partial \dot{q}^{\,i}_C/\partial q^i_C\right)$ associated with the coarse partition at such an equilibrium is strictly positive. Consequently (by Proposition 2) no such stable point can have $q^{*i}_C > 0$ for any $i = 1, 2$. Consider now restpoints at which the fine partition is used with asymptotic probability 1 by both players. 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 phenotypic play of the opponent in both games γ_1 and γ_2. As the coarse partition has smaller reasoning cost, the diagonal element of the Jacobian matrix associated with the linearization of the dynamics at this restpoint, $\left(\partial \dot{q}_C/\partial q_C\right) = \Xi(2) - \Xi(1) - 2\varepsilon_1 > 0$. The strict Nash equilibrium $\left(\sigma^1_{H1}, \sigma^2_{H1}\right) = (1,1)$ cannot be induced at any stable restpoint. ■

Proof of Proposition 5:

Proof. (i) Again as $\operatorname{card}\Gamma = 1$ there is trivially only one partition and one analogy class $g = \gamma_1$, where γ_1 is given by (27). The spectrum of the Jacobian matrix $\mathcal{M}$ associated with the linearization of the dynamics 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 $\{\lambda_1, \lambda_2\} = -2\varepsilon_0 \pm \sigma^i_{H1}(1 - \sigma^i_{H1})(a_1 + a_4 - (a_2 + a_3))$. Consequently $\mathcal{M}$ has an eigenvalue with $\lim_{\varepsilon_0 \to 0} \lambda_i(\varepsilon_0) > 0$ and $\widehat{\sigma}_1$ is unstable.

(ii) Let the 2×2 game γ_2 be again the game described in (27). 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 equilib-

rium of the strategic form (28) 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 $p^i_{HC} = \frac{a_4-a_3}{a_1+a_4-(a_2+a_3)}$ with asymptotic probability one. This restpoint is asymptotically stable whenever $f_1/f_2 < \widehat{f}(\Gamma)$ as can be shown in analogy to part (ii) of 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. On the other hand 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. ■

C Appendix: Proofs from Section 6

In the proof of Proposition 7 we will use the (negative) entropy function $\varphi(q) = -\sum_{\mathcal{G}} q_l \ln q_l$ as noise function (analogously for action choice frequencies). Using this function corresponds to using stochastic perturbations with extreme value distributions and leads to the logit choice function.[49] In general the results obtain with any function $\varphi(q)$ ($\varphi(p)$) satisfying the following: (i) $\varphi(q)$ should be strictly concave (i.e. $\varphi''(q)$ negative definite) and (ii) the gradient of $\varphi(q)$ should become arbitrarily large near the boundary of the phase space. See Hofbauer and Hopkins (2005).

Proof of Proposition 7:

Proof. (i) The first-order conditions for problem (12) are $\overline{\Pi}_l(\cdot) + \varepsilon_1 \varphi'(q_l) = 0, \forall G_l \in \mathcal{G}$ and $\sum_{G_l \in \mathcal{G}} q_l = 1$. With entropy as noise function the agent's choices are given by

$$q^{it}_l = \frac{\exp^{\varepsilon_1^{-1}\overline{\Pi}^{it}_l(\cdot)}}{\sum_{G_h \in \mathcal{G}} \exp^{\varepsilon_1^{-1}\overline{\Pi}^{it}_h(\cdot)}} =: B_l(\cdot) \tag{29}$$

Consider next the expected motion of the partition choice frequencies. We can write $\left\langle q^{i(t+1)}_l - q^{it}_l \right\rangle = B_l(\overline{\Pi}^{i(t+1)}(\cdot)) - B_l(\overline{\Pi}^{it}(\cdot))$ which can be approximated by $\left\langle q^{i(t+1)}_l - q^{it}_l \right\rangle = \sum_{G_h \in \mathcal{G}} \frac{\partial B_l(\cdot)}{\partial \overline{\Pi}^i_h(\cdot)} \left\langle \overline{\Pi}^{i(t+1)}_h(\cdot) - \overline{\Pi}^{it}_h(\cdot) \right\rangle + O\left(\frac{1}{t^2}\right)$.Noting that $\frac{\partial B_l(\cdot)}{\partial \overline{\Pi}^i_l(\cdot)} = \varepsilon_1^{-1} q_l(1-q_l)$ and $\frac{\partial B_l(\cdot)}{\partial \overline{\Pi}^i_h(\cdot)} = -\varepsilon_1^{-1} q_l q_h, \forall h \neq l$ we can rewrite the previous

[49] This function has been widely used in the literature. See Fudenberg and Levine (1998), Hopkins (2002) or Hofbauer and Sandholm (2002) among others.

equation as

$$\left\langle q_l^{i(t+1)} - q_l^{it} \right\rangle = \varepsilon_1^{-1} q_l \begin{bmatrix} (1-q_l)\left\langle \overline{\Pi}_l^{i(t+1)}(\cdot) - \overline{\Pi}_l^{it}(\cdot) \right\rangle \\ -\sum_{G_h \neq G_l} q_h \left\langle \overline{\Pi}_h^{i(t+1)}(\cdot) - \overline{\Pi}_h^{it}(\cdot) \right\rangle \end{bmatrix} + O\left(\frac{1}{t^2}\right).$$

Next note that $\left\langle \overline{\Pi}_l^{i(t+1)}(\cdot) - \overline{\Pi}_l^{it}(\cdot) \right\rangle = \frac{1}{t+\mu+1}[\Pi_l(\cdot) - \overline{\Pi}_l^{it}(\cdot)]$, where μ here is the weight placed on the initial beliefs. Furthermore it follows from the first-order conditions that $-\overline{\Pi}_l(\cdot) + \sum_{G_h \in \mathcal{G}} q_h \overline{\Pi}_h(\cdot) = \varphi(q) - \ln q_l =: \chi(q)$. Denoting $\varepsilon_1^{-1} = \kappa$, we have $\left\langle q_l^{i(t+1)} - q_l^{it} \right\rangle = \frac{\kappa}{t+\mu+1} q_l^t \left[S_l^{it} + \varepsilon_1 \chi(q) \right] + O\left(\frac{1}{t^2}\right)$ .The stochastic approximation then yields $\dot{q}_l^i = \kappa \left[q_l^i S_l^i(x) + \varepsilon_1 \chi(q) \right]$, which (up to a difference in the noise term and a multiplicative constant (κ)) is identical to (9).[50] Note though that the noise term $(\varepsilon_1 \chi(q))$ is still decreasing in q_l. The first-order conditions for problem (13) are given by $\sum_{\gamma_j \in g_k^i} f_j \sum_{a_n \in A_2} \pi(a_m^i, a_n^{-it}, \gamma_j) z_n^{-it}(g_k) + \varepsilon_0(1 + \ln p_{mk}^{it}) = 0$,and $\sum_{A_i} p_{mk}^i = 1$, $\forall a_m^i \in A_i, g_k \in \mathcal{P}^+(\Gamma)$. Denote by $B_m(\cdot)$ the associated choice functions and let

$E_z\left[\Pi_{mk}^{it}(\cdot)\right] = \sum_{\gamma_j \in g_k^i} f_j \sum_{a_n \in A_2} \pi(a_m^i, a_n^{-it}, \gamma_j) z_n^{-it}(g_k)$ be the expected payoff of player i when choosing action a_m given beliefs $z^{-it}(g_k^i)$. Note that $\frac{\partial B_m(\cdot)}{\partial E_z[\Pi_{mk}(\cdot)]} = \varepsilon_0^{-1} p_{mk}(1 - p_{mk})$ and $\frac{\partial B_m(\cdot)}{\partial E_z[\Pi_{hk}(\cdot)]} = -\varepsilon_0 p_{hk} p_{mk}$. Furthermore $\left\langle z^{-i(t+1)}(g_k^i) \right\rangle - \left\langle z^{-it}(g_k^i) \right\rangle = r_k^{it} \sum_{\gamma_j \in g_k} f_j \frac{\sigma_j^{-it} - z^{-it}(g_k)}{t+\mu}$. But then again we can write

$$\left\langle p_{mk}^{i(t+1)} - p_{mk}^{it} \right\rangle = \frac{\varepsilon_0^{-1}}{t+\mu} p_{mk}^{it} \begin{bmatrix} (1 - p_{mk}^{it}) \left[r_k^{it} \Pi_{mk}^{it}(\cdot) - E_z \Pi_{mk}^{it}(\cdot) \right] \\ -\sum_{a_h \neq a_m} p_{hk}^{it} \left[r_k^{it} \Pi_{hk}^{it}(\cdot) - E_z \Pi_{hk}^{it}(\cdot) \right] \end{bmatrix} + O\left(\frac{1}{t^2}\right).$$

From the first-order condition we get $\left\langle p_{mk}^{i(t+1)} - p_{mk}^{it} \right\rangle = \frac{\kappa}{t+\mu}[p_{mk}^{it} r_k^{it} S_{mk}^{it} + \varepsilon_0 \chi(p_k)]$ $+ O\left(\frac{1}{t^2}\right)$and thus $\dot{p}_{mk} = \kappa \left[p_{mk}^i r_k^i S_{mk}^i + \varepsilon_0 \chi(p_k) \right]$, which is identical to (8) up to a difference in the noise term (and a multiplicative constant). As $\chi'(p_k) < 0$ and furthermore the sign of $O(\varepsilon_0)$ is preserved the stability properties of the process are those of (8) - (9).

(ii) Now consider the process where agents correlate action and partition choice employing choice rule (14). The first-order conditions for problem (14) are given by $\sum q_l^{it} \mathrm{I}_{kl} \sum_{\gamma_j \in g_k^i} f_j \sum_{a_n \in A_2} \pi(a_m^i, a_n^{-it}, \gamma_j) z_n^{-it}(g_k) + \varepsilon \varphi'(p_{mk}) = 0, \forall a_m^i \in A_i, g_k \in \mathcal{P}^+(\Gamma)$ and $\sum_{a_m \in A_i} p_{mk}^i = 1$ as well as $\sum_{g_k \in G_l} p_k^{it} \; f_j \pi(\gamma_j)$ $z^{-it}(g_k^i) \mathrm{I}_{jk} + \varepsilon \varphi'(q_l) = 0, \forall G_l \in \mathcal{G}$ and $\sum_{G_l \in \mathcal{G}} q_l = 1$. These first-order conditions lead to the same choice functions where in (29) $\Pi_l^{it}(x^t)$ is used instead of the historical payoffs $\overline{\Pi}_l^{it}(\cdot)$. The stochastic approximation under choice rule (14) will coincide up to a multiplicative constant with that of rule (12)-(13). ■

Proof of Proposition 8:

[50] For the fictitious play process it is convenient to replace the assumption of vanishing noise by an assumption that $\varepsilon_1 = \varepsilon_0 = \varepsilon$ is a fixed but arbitrarily small number. In particular ε has to be smaller than the smallest increment of the reasoning cost function.

Proof. It follows immediately from the argument developed in the proof of Proposition 2 above that this proposition continues to hold. Furthermore whenever $\operatorname{card}\Gamma = 1$ the process SFP1 with choice rules (12) and (13) coincides with the process SFP2 with choice rules (12) and (15). Consequently part (i) of Propositions 3, 4 and 5 continues to hold. On the other hand some asymptotically stable restpoints under SFP1 will be stable under SFP2 only if additional conditions are met. Consider for example Proposition 3. A necessary condition for the equilibrium in weakly dominated strategies a_w^i to be phenotypically induced in γ_1 at a stable restpoint is that $a_w^{-i} \in BR(f_1 a^{*i} \oplus f_2 a_w^i | \gamma_1)$. The following example demonstrates that part (ii) of Proposition 3 can fail. Let two games occurring with the same frequency be given by

γ_1 :

	H	L
H	2, 2	3, 1
L	2, 1	4, 4

, γ_2 :

	H	L
H	5, 5	1, 0
L	0, 0	0, 0

.

(H, H) is a Nash equilibrium in weakly dominated strategies in γ_1 and a strict Nash equilibrium in γ_2. Now note that if player 1 (the row player) chooses L in γ_1 and H in γ_2, the best response of player 2 in game γ_1 to the belief $\frac{1}{2}H \oplus \frac{1}{2}L$ is to play L in game γ_1. Consequently $a_w^{-i} \notin BR(f_1 a^{*i} \oplus f_2 a_w^i | \gamma_1)$. Now starting from (H, H) a deviation by player 2 (to play strategy $\eta L \oplus (1-\eta)H$ for some small $\eta > 0$ in γ_1) immediately induces player 1 to play strategy L in γ_1 as a best response to this observation. But then in turn player 2 will choose L as a best response to the belief $\frac{1}{2}H \oplus \frac{1}{2}L$. Similar considerations are true for Propositions 4 and 5. The result relating choice rule (12) and (15) to choice rule (16) is shown in analogy to the proof of Proposition 7. ■