Learning Across Games*

Friederike Mengel†
University of Alicante
& Maastricht University

August 2008

Abstract

This paper studies the learning process carried out by two agents who are involved in many games. As distinguishing all games can be too costly (require too much reasoning resources) agents might 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. Learning across games can destabilize strict Nash equilibria and stabilize equilibria in weakly dominated strategies as well as mixed equilibria in 2×2 Coordination games even for arbitrarily small reasoning costs. The model is also able to explain experimental findings from e.g. the Traveler's dilemma or deviations from subgame perfection in Bargaining Games.

JEL Classification: C70, C72, C73.

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

---

*This paper has benefitted enormously from discussions with my supervisor Fernando Vega Redondo. I also wish to thank Ana B. Ania, Larry Blume, Jayant Ganguli, Veronika Grimm, Ani Guerdjikova, Ed Hopkins, Christoph Kuzmics, Fabrice Le Lec, Stef Proost, Karl Schlag, Jakub Steiner as well as seminar participants at Alicante, Bocconi, Budapest (ESEM 2007), California-Irvine, Cologne, Cornell, Edinburgh, Granada (SAE 2007), Hamburg (SMYE 2007), Heidelberg, IAE (Barcelona), Lausanne, Leuven, Maastricht, Madrid, Microsoft, Pompeu Fabra, Santa Fe Institute, Stony Brook, Urbana-Champaign, Warwick (RES 2007) and Wien for helpful comments. Financial support from the Spanish Ministery of Education and Science (grant SEJ 2004-02172) is gratefully acknowledged.

†Department of Economics (AE1), Faculty of Economics and Business, Maastricht University, P.O Box 616, 6200 MD Maastricht, Netherlands. *e-mail:* F.Mengel@algec.unimaas.nl

1 Introduction

Economic agents are involved in many games. Some will 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. Distinguishing all games at all times requires a huge amount of alertness or reasoning effort. Consequently it is natural to assume that agents might partition the set of all games into analogy classes, i.e. subsets 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 the 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 the different interactions colleagues face at work (e.g. in an Economics department). Focus on interactions where each player has two actions available, say to provide either high or low effort. Some of these interactions might correspond to coordination games, where it is in each player's best interest to match the action of the other. Others might correspond to Anti-coordination games, where each player's best response is to choose the action the opponent does *not* choose. And still others might correspond to games of conflict, where one player wants to match the action of the opponent, but the other player does not. If agents are involved in many such interactions, it is not clear whether they will want to distinguish all of them at all times. Doing so will certainly require a high amount of alertness (or a high reasoning cost). In fact we show in this paper, that (under some conditions on the frequencies with which these games occur) agents will not distinguish them, even for arbitrarily small reasoning costs. Furthermore in this case, the strict Nash equilibria in the Coordination Games will never be observed. This conclusion is our starting point to study the implications of learning across games for equilibrium selection in two-player games. Then, even for arbitrarily small reasoning costs, we establish the following results.

- Learning across games (if it converges) leads to approximate Nash equilibrium play in all games.[1]
- Nash equilibria in weakly dominated strategies that are unstable to learning in a single game can be stabilized by learning across games.

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

- Strict Nash equilibria that are always stable to learning in a single game can be destabilized by learning across games.
- Mixed Nash equilibria in 2×2 Coordination games that are unstable to learning in a single game can be stabilized by learning across games.

Furthermore we show that learning across games can explain experimental results such as those described by Goeree and Holt (2001) in the paper "Ten Little Treasures of Game Theory and Ten Intuitive Contradictions", but also deviations from subgame perfection sometimes observed in experiments with bargaining games. We also 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 to the use of alternative learning models and relate our results to the concept of analogy based expectations equilibrium introduced by Jehiel (2005).

The paper is organized as follows. In Section 2 the model is presented. In Section 3 we characterize equilibrium actions and partitions. In section 4 we show that learning across games can provide a natural explanation for several experimental results. Section 5 discusses alternative settings. In Section 6 we discuss related literature and Section 7 concludes. The proofs are relegated to an appendix.

2 The Model

2.1 Games, Partitions, Payoffs

Games and Partitions

There are 2 players indexed $i = 1, 2$ playing at each point in time $t = 1, 2...$ a strategic form game randomly and independently drawn from the set $\Gamma = \{\gamma_{1,}....\gamma_J\}$ according to probabilities $f = (f_j)_{\gamma_j \in \Gamma}$ where $f_j > 0, \forall \gamma_j \in \Gamma$.[2] For both players $i = 1, 2$ all games $\gamma \in \Gamma$ share the same action set A_i. Denote by $\mathcal{P}(\Gamma)$ the power set (or set of subsets) of Γ and $\mathcal{P}^+(\Gamma)$ the set $\mathcal{P}(\Gamma) - \varnothing$. Players partition the set of all games into subsets of games they see as analogous. Denote G a *partition* of Γ with $\text{card}(G) = Z$. An element $g \in G$ is called an *analogy class.* Analogy classes are denoted $g_k \in \mathcal{P}^+(\Gamma) = \{g_1, ..., g_K\}$. The set of all possible partitions of Γ is given by $\mathcal{G} = \{G_1, ..., G_L\}$ with $\text{card}(\mathcal{G}) = L$. Furthermore denote actions for player i by $a^i_m \in A_i = \{a^i_1, ..., a^i_{M(i)}\}$, where $M(i) = \text{card}\, A_i$. 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.

Payoffs and Reasoning Costs

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 $a^t = (a^{1t}, a^{2t})$.

[2] In the following we will denote both the random variable and its realization by γ.

In addition there is a cost $\Xi(Z,\xi)$ of distinguishing games reflecting the agents' limited reasoning resources. $\Xi(Z,\xi)$ is an increasing function of Z, i.e. partitions of higher cardinality are more costly. More precisely: $Z_l \gtreqless Z_h \Leftrightarrow \Xi(Z_l,\xi) \gtreqless \Xi(Z_h,\xi)$. The parameter ξ gives an upper bound on reasoning costs, i.e. $\forall\ Z, \xi > 0$: $0 \leq \Xi(Z,\xi) < \xi$.

Note that in reality reasoning costs can also apply to memorizing larger quantities of information, employing more complex strategies or to learning in a larger state space etc... We assume (as in standard models) that all these activities are costless. The only activity that we assume has a cost is distinguishing different games. This is done in order to minimize the number of deviations from standard models.[3] Finally we assume that net payoffs (net of reasoning costs) are positive and finite.[4]

2.2 Learning

The learning model we consider is one of reinforcement 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 $g_k \in \mathcal{P}^+(\Gamma)$. 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. After playing a game players update their propensities and attractions taking into account the payoff obtained.[6]

At any point in time t a player is thus 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 (depending on the analogy class).

The Dynamic Process

The dynamic process unfolds as follows.

(i) First players choose a partition G_l with probability

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

[3] In Section 5.3 we will discuss our assumptions on reasoning costs some more.

[4] Normalizing payoffs to be positive is a technical assumption commonly used in reinforcement models. (See among others Börgers and Sarin (1997)). Here of course it implies that reasoning costs have to be "small" compared to the game payoffs. This is the most interesting case, though.

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

[6] While reinforcement learning as applied to action choices has been extensively studied in the literature and has strong empirical support, learning of partitions (or categorizations) is not very well understood. A priori, though, no sufficient reason seems to exist to justify why agents should learn about partitions in a different way than they do about actions. Very few papers on reinforcement learning of categorizations exist in the computer science literature. See Baum (2004) or Porta (1999) among others.

Denote G^{it} the partition actually chosen by player i at time t.[7]

(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}. The reasoning cost $\Xi(Z(t), \xi)$ is incurred.

(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 fixed amount $\varepsilon_0 > 0$. In the analogy class just visited ε_0 is best interpreted as noise or experimentation.[8] 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.

(vi) Players update propensities as follows:

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

where again $\varepsilon_1 > 0$ is noise. The payoffs relevant for partition updating are payoffs net of costs of holding partitions.[9]

Action Choice and Phenotypic Play

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 $p_k^{it} = (p_{1k}^{it}, ..., p_{Mk}^{it})$. 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 phenotypic play probability σ_{mj}^{it} captures the overall probability (across partitions) with which action m is

[7] One might think that agents do not choose a new partition each time they face a new game, but more infrequenty. As such a change only affects the relative speed of learning it doesn't affect our results qualitatively.

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

[9] One might wonder why agents do not update *all* partitions that contain g^{it}. Such "cross-updating" would be an additional deviation from standard reinforcement learning in a single game. We want to focus in our analysis though on the isolated effect of the cost of distinguishing different games.

chosen when the game is γ_j. It is generated from 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.

Flat Learning Curves and Step Size

A 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 decreasing step sizes greatly simplifies the study of the asymptotic behavior of the process as we will illustrate now.

2.3 Stochastic Approximation and an Evolutionary Interpretation

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.

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.[10] 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 $\sum_{t=1}^{\infty} \left(\frac{1}{\beta_k^{it}}\right)^2 < \infty, \forall g_k \in \mathcal{P}^+(\Gamma), i = 1,2)$ described above. 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. One possibility 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.[11]

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 $\alpha^{it} = \mu + t\theta$

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

[11] 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.

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.[12] Then there is a unique step size of order t^{-1}. Call the resulting process the *normalized* process.

Now denote by $\Pi_{mk}^{it}(x^t)$ the expected payoff of action m conditional on visiting analogy class g_k (for player i at time t). 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: $S_{mk}^{it}(x^t) = \Pi_{mk}^{it}(x^t) - \sum_{a_h \in A_i} p_{hk}^{it} \Pi_{hk}^{it}(x^t)$. Analogously define $S_l^{it}(x^t)$ for partition G_l and let r_k^{it} be the total frequency with which analogy class g_k is visited at t.

Stochastic Approximation Theory now shows that (under the conditions mentioned) a discrete time stochastic process can be approximated by its mean ODE (that we describe in the following Proposition). The behavior of this mean ODE and the stochastic process are very closely linked, as we will state precisely right after the statement of Proposition 1.

Proposition 1 *The system of mean ODE's approximating our (normalized) stochastic learning process is*

$$\dot{p}_{mk}^{i} = p_{mk}^i r_k^i S_{mk}^i(x) + \varepsilon_0(1 - M p_{mk}^i), \forall a_m \in A_i, g_k \in \mathcal{P}^+(\Gamma) \tag{5}$$

and

$$\dot{q}_l^{i} = q_l^i S_l^i(x) + \varepsilon_1(1 - L q_l^i), \forall G_l \in \mathcal{G}, i = 1, 2. \tag{6}$$

Proof. Appendix B. ■

The evolution of the choice probabilities $x^{it} = (p^{it}, q^{it})$ is closely related to the behavior of the deterministic system (5)-(6). More precisely let us denote the vector field associated with the system (5)-(6) 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 $F(x(t))$. 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 $\Pr\{\lim_{t\to\infty} x^t = x^*\} > 0$.[13] In the following analysis we will thus focus on the asymptotically stable points of (5)-(6).

An Evolutionary Interpretation

One interpretation of our model is thus that ultimately evolution selects among partitions and actions. More precisely we have in mind an evolutionary model with two large populations whose members are randomly matched in pairs to play a game drawn from the set Γ according to the frequencies defined

[12] 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.

[13] See Benaïm and Hirsch (1999), Benaïm and Weibull (2003), Benveniste, Métivier and Priouret (1987), Kushner and Lin (2003) or Pemantle (1990). For the reader not acquainted with dynamical systems we define some of the terminology in the appendix.

above. Each player is genetically programmed to use a certain partition and a profile of actions measurable with respect to this partition. In some abuse of notation denote the two populations by $i = 1, 2$. Then we can think of selection being determined by the set of equations (5)-(6).

The dynamics (5)-(6) is a perturbed version of the well known replicator dynamics. The replicator dynamics is the canonical model of evolution in Evolutionary Game Theory, but also Biology (see e.g. the textbooks by Weibull (1995) or Vega-Redondo (2000)). It has been extensively studied in the literature and how it selects among different equilibria in a single game is well understood. The relation between perturbed reinforcement learning and replicator dynamics has been pointed out by Hopkins (2002). (Note also that if $\operatorname{card}\Gamma = 1$, there is obviously only one partition and the above system reduces to the standard replicator dynamics in a single game, which is described e.g. in Chapter 5.2 of Weibull (1995)).

Throughout the analysis, we will assume that noise is vanishingly small and of the same order for both action and partition choices.

Assumption A1: (i) $\varepsilon_0 \to 0$ and (ii) $\varepsilon_1 = \lambda\varepsilon_0$ for some constant λ.

We explain in detail why we make this Assumption at the beginning of Appendix C.

3 Results

In this section we will present a set of more general results for the case of arbitrarily small reasoning costs. The results in this section are of a more theoretical interests. They show that there is an interesting discontinuity in the sense that even for arbitrarily small reasoning costs, the selection of learning across games can be very different from the selection in the case where there are no reasoning costs at all. In section 4 we will then proceed to show how learning across games can provide a natural explanation for several well known experimental results.

3.1 Action Choices

Our first result 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 $\overline{\xi}(\Gamma, f) > 0$ s.t. whenever $\xi < \overline{\xi}(\Gamma, f)$ 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 C. ■

Whenever reasoning costs are small enough equilibrium action and partition choices will be such that approximately a Nash equilibrium is played in *each* of the games. This is the case even if players do not distinguish between games. Thus - unless reasoning costs are significant - learning across games does not lead to deviations from this basic prediction of game theory.[14]

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

3.1.1 Two examples

Let us start out with two intuitive examples. Consider the following set of three games $\Gamma_1 = \{\gamma_1, \gamma_2, \gamma_3\}$.

$$\gamma_1 : \begin{pmatrix} 2,2 & 1,1 \\ 1,1 & 2,2 \end{pmatrix}, \gamma_2 : \begin{pmatrix} 1,1 & 2,2 \\ 2,2 & 1,1 \end{pmatrix}, \gamma_3 : \begin{pmatrix} 2,1 & 1,2 \\ 1,2 & 2,1 \end{pmatrix}$$

γ_1 is a game of pure Coordination, γ_2 is an (Anti)-Coordination Game and 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). 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.[15] Simultaneous learning of actions and partitions leads to the same prediction whenever there are no reasoning costs ($\Xi(Z, \xi) = 0, \forall Z \in \mathbb{N}$). For arbitrarily small costs things change. This is the discontinuity mentioned above. 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 1 *Assume $f_j < 1/2, \forall j = 1, 2$. Then $\forall \xi > 0$ the unique asymptotically stable point for Γ_1 involves both players holding the coarsest partition $G_C = \{\gamma_1, \gamma_2, \gamma_3\}$ with asymptotic probability 1 and choosing $p_c^* = 1/2$.*

Proof. Appendix C. ■

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. Note though that both games

[14] 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.

[15] This is shown in Appendix C.

together can have probability f_1+f_2 very close to one. If and only if $f_j<1/2$ for $j=1,2$ the incentives of an agent who sees the three games as "one" correspond to those of a conflict game. Consequently, playing the mixed equilibrium with the coarse partition is asymptotically stable under this condition. Note also that - as the equilibrium in Claim 1 is unique - the presence of the conflict game destabilizes the otherwise asymptotically stable strict equilibria in the Coordination Games.

Not always does this require that agents won't distinguish the games in equilibrium. Consider the following example of two games $\Gamma_2=\{\gamma_1,\gamma_2\}$,

γ_1:

	a	b	c
A	$\mathbf{2,2}$	$1,1$	$2,1$
B	$1,2$	$2,2$	$5,7$
C	$1,1$	$7,5$	$\mathbf{8,6}$

; γ_2:

	a	b	c
A	$6,4$	$4,6$	$3,3$
B	$4,6$	$6,4$	$3,3$
C	$3,3$	$3,3$	$2,2$

.

Game 1 has two strict Nash equilibria: (A,a) and (C,c). Both are stable to learning in a single game. By contrast learning across games singles out (C,c) as a unique prediction for γ_1. More precisely we can state the following claim,

Claim 2 *Assume* $1>f_1>\frac{-17+\sqrt{409}}{6}$. *Then there exists* $\overline{\xi}(\Gamma_2)>0$ s.t. $\forall\xi\in(0,\overline{\xi}(\Gamma_2))$ *the unique as. stable point* x^* *has both players holding the finest partition* $G_F=\{\{\gamma_1\},\{\gamma_2\}\}$ *and play* (C,c) *in* γ_1 *and* $(\frac{1}{2}A\oplus\frac{1}{2}B\ ,\frac{1}{2}a\oplus\frac{1}{2}b)$ *in* γ_2 *with asymptotic prob.* 1.

Proof. Appendix C. ■

If the first game occurs sufficiently often, then it is impossible to induce the strict Nash equilibrium (A,a) at an asymptotically stable point. The reason is that A is a best response to the mixed equilibrium $(\frac{1}{2}A\oplus\frac{1}{2}B\ ,\frac{1}{2}a\oplus\frac{1}{2}b)$ that will be observed with probability 1 in game γ_2 (by Proposition 2). This induces a tendency not to distinguish between the two games in order to save reasoning costs, destabilizing the strict Nash equilibrium (A,a).[16] It can be shown that points involving the coarse partition are not stable either. At the unique stable point agents will thus distinguish the two games. Nevertheless the predictions from learning across games differ from those of learning in a single game in that (C,c) is the *unique* prediction in game γ_1. In this example learning across games thus does not only lead to different, but also stronger predictions than learning in a single game, predicting that (C,c) will always be observed in game γ_1. In that sense learning across games can lead to both "positive" and "negative" new results. "Positive", as the model can have more predictive power than learning in a single game, and "negative", as strict Nash equilibria can be destabilized.

We have seen that some strict Nash equilibria will never be observed as an outcome of the learning across games process irrespective of which partition agents will hold in equilibrium. In other cases agents might even end up mixing

[16] For high reasoning costs ($\xi>\overline{\xi}(\Gamma_2)$) the strict Nash equilibrium (A,a) cannot be induced either whenever $f_1<\frac{2}{3}$.

between different partitions at an asymptotically stable point, but unfortunately convergence is not always guaranteed.

3.1.2 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.[17] Learning across games can sometimes destabilize strict Nash equilibria as we have seen in the previous examples. This point is made more precise and general in the following proposition.

Proposition 3 *Let* $\widehat{\sigma}_1 = (\widehat{\sigma}_1^1, \widehat{\sigma}_1^2)$ *be a strict Nash equilibrium in* $\gamma_1 \in \Gamma$ *. 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 unique equ. in mixed strategies stable to learning in a single game with* $\widehat{\sigma}_1$ *in its support. Then there exists* $\overline{f}(\Gamma) < 1$ and $\underline{f}(\Gamma) > 0$ *s.t. if* $f_1/f_2 \in (\underline{f}(\Gamma), \overline{f}(\Gamma))$ *the strict equ.* $\widehat{\sigma}_1$ *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. This is a standard result. 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. For arbitrarily small reasoning costs (provided that they are more important than noise) there will be a tendency for agents to see the games as analogous and to save reasoning costs.

As we have seen in the previous section (3.1.1), this can have "negative" or "positive" implications in terms of selection. On the one hand the predictive power in a particular game can be increased, on the other hand a fundamental result in learning in a single game (namely the stability of strict Nash equilibria) is shown not to hold.

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

3.1.3 Mixed Strategy Nash Equilibria in Coordination Games

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

Proposition 4 *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.[18] 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.

3.1.4 Stabilization of Nash equilibria in weakly dominated strategies

Finally we will show that Nash equilibria in weakly dominated strategis can be stabilized by learning across 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 5 *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 $x^* \in E^*$.*

[18] 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) for recent research on this topic.

(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 C. ■

Interesting implications of Proposition 5 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.[19] We will see how this proposition plays out in Section 4.3 in an application to bargaining 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.

3.2 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 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.[20] Now as an 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.t. 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* $S^{Nash}(\gamma_j) \cap S^{Nash}(\gamma_h)$

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

[20] Rubinstein (1988) uses distance in payoff (or probability) space as a similarity criterium in one-person decision problems. Steiner and Stewart (2008) use such a criterium for games.

$= \varnothing$, $\forall \gamma_j \neq \gamma_h \in \Gamma$. *Furthermore in this case the conclusions of part (i) of Propositions 3, 4 and 5 hold true in each of the games.*

Proof. Appendix C. ■

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 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.

4 Learning Across Games and Experiments

In this section we will not always assume that reasoning costs are *arbitrarily* small, but we will maintain the assumption that they are "small" in the sense that $\xi < \left|\min_{A_1 \times A_2 \times \Gamma} \pi^i(a, \gamma)\right|$, i.e. that they are always smaller than the smallest possible payoff in any of the games. Then we will show how learning across games can provide an intuitive explanation for many typical experimental results.

The experimental results we consider are largely taken from a paper by Goeree and Holt (2001), entitled "Ten Little Treasures of Game Theory and Ten Intuitive Contradictions". In this paper they report results on ten experimental games. Each game comes in two or three slightly different versions (with slight changes in payoffs) that yield dramatically different behavior in the laboratory. We try to argue in this section that these different results have a very natural explanation in terms of learning across games.

All the games in their experiments are played only once in the laboratory and different games are played by different participants. There are thus no opportunities for learning across games *in* the laboratory. This is true of almost all laboratory experiments. What we argue is that experimental subjects when they come to the lab and face an artificial situation they draw analogies to situations they have faced *outside* the lab. Goeree and Holt (2001) themselves remark in the conclusions: "The decision makers[...], like the subjects in our experiments, typically have experience in similar games with other people". We will thus posit that a process of learning across games takes place *outside* the laboratory and that subjects *inside* lab play according to the analogy class that is triggered by the game they face in the lab.

In the next subsections we will describe two of Goeree and Holts' (2001) puzzles (namely the Traveler's Dilemma and a type of Coordination games called "Kreps games"), that we think have a very intuitive explanation in terms of learning across games. We will then show that learning across games can also explain deviations from subgame perfection often observed in experimental bargaining games. In the Appendix we give an informal account of how other results in Goeree and Holt (2001) might be explained through learning across games.

4.1 The Traveler's Dilemma

The story in the Traveler's dilemma goes as follows. Two travelers have purchased identical antiques while on a tropical vacation. Their luggage is lost on the return trip and the airline asks them to make independent claims for compensation. The airline representative announces: "We know that the bags have identical contents, and we will entertain any claim between 180$ and 300$, but you will be reimbursed only the minimum of your two claims. In addition - if your claims differ - an amount of R will be transferred from the one making the higher claim to the one making the lower claim". The action set is $A = [180, 300] \cap \mathbb{N}$. Now if $R = 0$ the traveler's have nothing to lose by making the highest claim (300$), which is in fact a weakly dominant strategy. (All points where players claim the same are Nash equilibria). As soon as $R > 1$ on the other hand claiming 180$ is the unique Nash equilibrium and the unique rationalizable strategy.

Goeree and Holt (2001) play two games in the lab. One where $R = 5$ and one where $R = 180$. They find that in the treatment where $R = 5$ a large majority of subjects choose $a_i = 300$, whereas in the treatment where $R = 180$ most subjects choose $a_i = 180$.[21]

Remark: When we try to explain the experimental findings we think it is more insightful to explain the predominant pattern found in the data, rather than predicting the exact distribution of choices. This could be done by playing with the noise parameter ε in the model, but could just as well be due to things not modeled. Some subjects might fail to map the experimental game into the "correct" analogy class or there might be differences in reasoning costs across subjects etc..

To explain these results we assume that *outside* the lab participants regularly face games with this payoff structure

$$\pi_i(a, R) = v + \min\{a_i, a_j\} - R * (-1)^{\delta(a_i < a_j)}$$

[21] More precisely in the treatment $R = 5$ (180) around 80% choose 300 (180) and around 5% choose 180 (300), 190 (270), 250 (240) and 280 (190). Only a graph with the distribution of choices is available in Goeree and Holt (2001).

where $\delta(a_i < a_j) = 1$ if $a_i < a_j$ and zero otherwise. The number $v \geq 2$ is to ensure that payoffs are always positive and can be interpreted as the "show up fee" (in their experiment 6$) that participants were paid just for arriving at the lab. The action set is $A = [180, 300] \cap \mathbb{N}$ and the set of games is defined by the different values R is allowed to take. For simplicity we assume that $R \in \{0, 5, 180\}$. It seems intuitive to us that many (if not a large majority) of games played in real life have the more simple structure where $R = 0$. If we define the set of games as $\Gamma_{TD} = \{\gamma_0, \gamma_5, \gamma_{180}\}$ we have 5 possible partitions. Denote by f_0, f_5 and f_{180} the frequencies with which these three games occur respectively. For clarity of exposition, assume that reasoning costs have the following simple structure: $\Xi(Z, \xi) = \widetilde{\xi} * (Z - 1)$ for some $\widetilde{\xi} > 0$. We can then state the following claim.

Claim 3 *If $f_5 < \widetilde{\xi}/4 < f_0/4$ there is an asymptotically stable point where both players hold partition $G^* = \{\{\gamma_{180}\}, \{\gamma_0, \gamma_5\}\}$ and choose* 180 *in analogy class $\{\gamma_{180}\}$ and* 300 *in analogy class $\{\gamma_0, \gamma_5\}$ with asy. prob. one.*

Proof. Appendix C. ■

Claim 3 implies a deviation from Nash equilibrium in the game with $R = 5$ in spite of the fact that reasoning costs need not be "large" compared to the payoffs of the game. They cannot be *arbitrarily* small though, as it was the case in Section 3, in order to have the above result. Note also that it is not the case that learning across games can explain "everything". The reverse result for example (where everyone chooses $a = 300$ in the game with $R = 180$ and $a = 180$ in the game with $R = 5$) could not be explained by learning across games, no matter how high the reasoning costs, how rarely the games occur or what other games are played.

4.2 Kreps Games

With this example (also taken from Goeree and Holt (2001)) we want to illustrate how learning across games can select between the conflicting motives of risk-dominance and payoff dominance in Coordination games . Another reason why we find the example of some independent interest is that it illustrates that players in different player roles need not end up with the same partitions at a stable point. The results we consider are from a set of games that Goeree and Holt (2001) call "Kreps games", which have the following payoff matrix.

	Left	Middle	NonNash	Right
Top	$500, 350$	$300, 345$	$310, 330$	$320, 50$
Bottom	$300, 50$	$310, 200$	$330, 330$	x, y

Clearly ("Top", "Left") is always a Nash equilibrium in this game. ("Bottom","Right") is a Nash equilibrium whenever $(x, y) \geq (320, 330)$ and ("Bottom", "Non-Nash") is a Nash equilibrium whenever $y \leq 330$.

Goeree and Holt (2001) report that if $(x, y) = (350, 340)$ a large majority of row players choose "Top" while roughly 2/3 of column players tend to choose

"Non-Nash" and 1/3 Left. If $(x, y) = (650, 700)$, though, almost all row players choose "Bottom" and almost all column players choose "Right".[22] How can we explain this result ? Clearly it seems to have something to do with the classical trade off between payoffs and risk. In particular most column players seem to choose the risk-free action "Non-Nash" in the first case, but not in the second. (For row players no strong trade-off exists between the two). To predict the behavior of the participants in this way though one has to make assumptions on their priors, degree of risk-aversion etc.. We want to illustrate now how learning across games how these motives emerge naturally from a process of learning across games.

Consider sets of games in which $x \in [300, 650] \cap \mathbb{N}$ and $y \in [50, 700] \cap \mathbb{N}$ (the maximum and minimum payoff in the matrix above). From this set we will consider three games for simplicity; the two games from the experiment and one game where ("Bottom","Non-Nash") is a Nash equilibrium. In particular we assume that in $\gamma_1 : (x_1, y_1) = (350, 340)$ in $\gamma_2 : (x_2, y_2) = (420, 320)$ and in $\gamma_3 : (x_3, y_3) = (650, 700)$, but any other choice for γ_2 s.t. $y_2 \leq 320$ would do as well. We assume again that reasoning costs have the following simple structure: $\Xi(Z, \xi) = \widetilde{\xi} * (Z - 1)$, where $\widetilde{\xi} > 0$. We also assume that $f_2 \geq \frac{1}{15} f_1$ and that there is some heterogeneity among players in reasoning costs.

Claim 4 *If $\widetilde{\xi} < 35 f_1$ there is an asymptotically stable point at which row players hold partition $G = \{\{\gamma_1\}, \{\gamma_2, \gamma_3\}\}$ and choose "Top" in γ_1 and "Bottom" in $\{\gamma_2, \gamma_3\}$. Column players hold partition $G_F = \{\{\gamma_1\}, \{\gamma_2\}, \{\gamma_3\}\}$ if $\widetilde{\xi} < 20 f_1$ and $G' = \{\{\gamma_1, \gamma_2\}, \{\gamma_3\}\}$ if $\widetilde{\xi} \in [20 f_1, 35 f_1)$. They choose "Left" in $\{\gamma_1\}$, "Non-Nash" in $\{\gamma_2\}$ and $\{\gamma_1, \gamma_2\}$ and "Right" in $\{\gamma_3\}$.*

Proof. Appendix C. ■

Of course the more heterogeneous behavior we want to explain the more heterogeneity we need in reasoning costs.[23] Note, though, that for all players reasoning costs can be very small compared to the payoffs of the game. Note also that learning across games could again not explain "everything" here. In particular it cannot explain the reverse choices, where players choose ("Bottom","Right") in γ_1 and Top and Non-Nash/Left in γ_3, unless one makes very contrived assumptions about the frequencies of the games. (In particular the ratio $\frac{f_2}{f_3}$ would need to be contained in the interval $[0.58, 0.71]$ in the most favorable case. As y_3 rises the size of this interval shrinks and for large enough y_3 this result could not be explained anymore.)

What explanation do Goeree and Holt (2001) offer for their results ? They argue that level k rationality theory by Stahl and Wilson (1995) can explain the

[22] More precisely if $(x, y) = (350, 340)$ 84% of row players choose "Top", 24% of column players "Left", 12% "Middle" and 64% "Non-Nash". If $(x, y) = (650, 700)$, 96% of row players choose "Bottom" and 84% of column players "Right", while 16% choose "Non-Nash". Note that we use the version with the "positive" payoff frame. The distribution of action choices is not substantially different, though, in the negative frame.

[23] It would also be possible to explain the results in terms of a mixed equilibrium where players are indifferent between coarser and finer partitions for a fixed reasoning cost - but this requires very precise assumptions on the parameters which we think are hard to justify.

data in conjunction with the logit choice rule with errors whenever a) higher levels of reasoning and b) a "higher complexity" of the game imply higher errors. (What "high complexity" means remains undefined in their paper). They also show that popular theories such as "inequity aversion theory" or "maxmin heuristics" cannot explain the data. We find it very natural to think that these results arise because the games trigger different analogy classes with the inexperienced players in the laboratory. We have seen that the model of learning across games can explain the data under weak and natural assumptions on the model parameters.

4.3 Bargaining

Finally we will also show how learning across games could explain deviations from subgame perfection often observed in experimental bargaining games. We do not use the bargaining games from Goeree and Holt (2001), as they only provide information on average behavior of the first proposers in their paper. Instead we consider a little more general set up, which is maybe also of independent interest to the reader. In the bargaining game two individuals, 1 and 2, 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. For simplicity we restrict the strategy space and focus on stationary strategies, i.e. strategies that do not depend on the decision node. A strategy of a player i is then characterized by two numbers (a^i, b^i) where a^i is the proposal (the share of pie she wants to keep) and b^i the acceptance threshold. Let a^i and b^i be from the finite grid $A = \{0, \frac{1}{M}, \frac{2}{M}, .. \frac{M-1}{M}, 1\}$, where M is some large but finite number. Finally for any number a denote by $\lfloor a \rfloor_{\frac{1}{M}}$ the largest integer multiple of $\frac{1}{M}$ that is smaller than a.

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 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.[24] Finally attractions and partitions are updated.

In particular let us consider the case where game γ_1 has discount factor $\delta_1 > 0$, and game γ_2 has discount factor $\delta_2 = 0$. Game γ_2 is essentially an Ultimatum Game, where the whole pie is gone if the first offer is not accepted. Both games have many Nash equilibria. As the action grid gets fine enough $(M \to \infty)$, all SPNE of the games tend to $a^i = \frac{1}{1+\delta_j}$ and $b^i = \frac{\delta_j}{1+\delta_j}$.[25] 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.

[24] Learning is thus based on the strategic forms in this example. For a model of learning in extensive form games see Laslier and Walliser (2005). They show that in generic perfect information extensive form games reinforcement learning within the game will lead to the subgame perfect Nash equilibrium. See also Jehiel and Samet (2005).

[25] See Propositions 3 and 4 in van Damme, Selten and Winter (1990).

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.

Claim 5 *$\forall \xi > 0$ there exists an asymptotically stable point for $\Gamma_3 = \{\gamma_1, \gamma_2\}$ involving both players holding the coarse partition $G = \{\gamma_1, \gamma_2\}$, player 1 demanding $a^{1*} = \left\lfloor \frac{1}{1+\delta_1 f_1} \right\rfloor_{\frac{1}{M}}$ and player 2 accepting all offers of at least $b^{2*} = 1 - \left\lfloor \frac{1}{1+\delta_1 f_1} \right\rfloor_{\frac{1}{M}}$ with asymptotic probability 1. If $\Xi(2) \geq \frac{f_1[1+\delta_1(1-f_1)]}{1+\delta_1}$ this is the unique stable point.*

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). 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. Any arbitrarily small reasoning cost suffices to stabilize the equilibrium with the coarse partition provided that it is more important than noise. There are many experiments that show that subjects often do not behave in accordance with subgame perfection.[26] 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 in these games.

5 Extensions and Discussion

There are some 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 contrast our results to previous results obtained by Jehiel (2005). And thirdly we would like to shortly discuss our assumptions on reasoning costs.

5.1 Other Learning Models

The other model (apart from reinforcement learning) that has received a lot of attention in the literature is the model of stochastic fictitious play.[27] In

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

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

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

$$z_m^{it}(g_k^{-i}) = \frac{\sum_{\tau=1}^{t-1} \delta_m^i(\tau)\delta_k^{-i}(\tau)}{\sum_{\tau=1}^{t-1} \delta_k^{-i}(\tau)} \tag{7}$$

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

$$\overline{\Pi}_l^{it}((x_\tau)_{\tau=1}^{t-1}, Z_l) = \frac{\sum_{\tau=1}^{t-1} \left[\pi^i(a^\tau, \gamma^\tau) - \Xi(Z_l)\right] \delta_l^i(\tau)}{\sum_{\tau=1}^{t-1} \delta_l^i(\tau)} \tag{8}$$

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) 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

$$q^{it*} \in \arg\max \sum_{G_h \in \mathcal{G}} q_h^t \overline{\Pi}_h(\cdot) + \varepsilon_1 \varphi(q^t) \tag{9}$$

where $\varphi(q)$ is a deterministic perturbation.[28] Some restrictions on the shape of this function are given in Appendix D. 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,

$$p_k^{it*} \in \arg\max p_k^{it} \left[\sum_{\gamma_j \in g_k^i} f_j \pi(\gamma_j)\right] z^{-it}(g_k^i) + \varepsilon_0 \varphi(p_k^t). \tag{10}$$

where $\pi(\gamma_j)$ is the payoff-matrix of game γ_j. Agents use the average payoff ma-

[28] Hofbauer and Sandholm (2002) have shown that for any stochastic perturbation used in (9) there is always an alternative representation using a deterministic noise function.

trix across all the games contained in analogy class g_k as relevant information.[29]

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

$$(q^{it*}, p_k^{it*}) \in \arg\max \sum_{G_h \in \mathcal{G}} q^{it} \left[\sum_{g_k \in G_l} p_k^{it} \left[\sum_{\gamma_j \in g_k^i} f_j \pi(\gamma_j) \right] z^{-it}(g_k^i) \right] \mathrm{I}_{jk} + \varepsilon\varphi(p^t, q^t), \tag{11}$$

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.[30]

Proposition 7 *Under stochastic fictitious play with choice rules (9) - (10) or (11) Propositions 2-6 as well as Claims 1-5 continue to hold.*

Proof. Appendix D. ■

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

5.2 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 his model players always know which game they are playing, but they do not distinguish between the play of the opponents in the different games. In the current paper on the other hand, players may not distinguish between games. In this section we use Jehiel's (2005) concept of analogy thinking and add an endogenous partition choice relying on stochastic fictitious play.

Then in the case of no correlation choice rule (10) is replaced by,

$$p_j^{it*}(g_k) \in \arg\max p_j^{it}(g_k)\pi(\gamma_j)z^{-it}(g_k^i) + \varepsilon_0\varphi(p_{jk}). \tag{12}$$

Note that the choice variable here is $p_j^{it}(g_k)$ instead of p_k^{it} in equation (10). 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.

[29] 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".

[30] 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.

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

$$(q^{it*}, p_j^{it*}(g_k)) \in \arg\max \sum_{G_h \in \mathcal{G}} q^{it} \sum_{g_k \in G_l} \sum_{\gamma_j \in \Gamma} f_j \left[p_j^{it}(g_k) \pi(\gamma_j) z^{-it}(g_k^i) \right] + \varepsilon \varphi(p^t, q^t). \tag{13}$$

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 (5) - (6). 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 (9) - (12) or (13) Proposition 2 continues to hold. On the other hand part (ii) of Propositions 3-5 will not hold in general.*

Proof. Appendix D. ■

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 (9) - (12) or (13) 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.

5.3 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. Also, one might argue that it is always optimal (in the sense of evolutionary selection) for an agent to have small reasoning costs. If this were the case one could argue that reasoning costs have evolved 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. 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.[31] Another question is whether much would change if one allows for alternative reasoning cost functions. The number of possibilities here is large.

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

Alternative Reasoning Cost-Functions

A natural alternative would be to assume that only *finer* partitions are more costly. A disadvantage of this assumption is that it does not deliver a complete ordering. On the other hand if there are only two games (like in many of our examples) the two notions coincide, as in this case partitions of higher cardinality are also finer. The essence of our results would thus be unaffected by such a change, while the exact statement of the propositions would certainly be different.

A different possibility could be that games have observable characteristics and that it is less costly to distinguish two games if they are "closer "in terms of this characteristic. While this does not seem unnatural, the problem is of course to choose these characteristics appropriately. One could think of the different situations or games coming with different "labels" and reasoning costs being determined how close the games are in terms of these labels. The problem with this assumption is that the choice of labels is ad hoc while any result one might hope to get will depend on how one chooses these labels.

One possibility would be to choose as a label some part of the description of the game, like e.g. payoffs. Costs would then be increasing in the distance in payoff space between two games (see also Section 4).[32] While this seems like a reasonable criterion in decision problems, it is much more problematic in games. (See section 3.2). Grimm and Mengel (2008) find in an experiment where participants play multiple games for 100 rounds that distance in payoff space cannot explain any of the observed behavior (according to different metrics considered). Another possibility would be to consider situations where games have many different labels and agents have to choose to which of these to pay attention. Binmore and Samuelson (2002) have investigated a related question. Such analysis, though, would be beyond the scope of the present paper.

6 Related Literature

The idea that similarities or analogies play an important role for economic decision making has long been present in the literature.[33] Most approaches have been axiomatic. Rubinstein (1988) gives an explanation of the Allais-paradox based on agents using similarity criteria in their decisions. Also Gilboa and Schmeidler (1995) argue that agents reason by drawing analogies to similar situations in the past. They derive representation theorems for an axiomatization of such a decision rule. 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

[32] This would entail an additional assumption of choosing the appropriate metric to measure this distance. (Candidates might be the Hausdorff distance between the convex hulls of payoffs or the distance between the matrix norms).

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

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 (2008) 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 bargaining 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 5 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.[34]

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

7 Conclusions

In this paper we have presented and analyzed a learning model in which decisionmakers learn simultaneously about actions and partitions of a set of games. We find that in equilibrium agents will partition the set of games according to strategic compatibility of the games. If the sets of Nash equilibria of any two games are disjoint agents will always distinguish these games in equilibrium. Whenever this is not the case though, interesting situations arise. In particular learning across games can destabilize strict Nash equilibria, stabilize Nash equilibria in weakly dominated strategies and mixed equilibria in 2×2 Coordination games.

Furthermore we have seen that learning across games can explain several experimental results, such as the Traveler's dilemma, but also deviations from subgame perfection that are sometimes observed in experiments. We find it very intuitive that participants that face a new situation in an experiment try to draw analogies to similar situations experienced outside the laboratory. We

[34] 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.

conjecture that analogy thinking could 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] Baum, E. (2004), "What is Thought", Cambridge, Mass: MIT Press.

[3] 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.

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

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

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

[7] Blume, L. (1993), The statistical mechanics of strategic interaction, Games and Economic Behavior, 5, 387-424.

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

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

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

[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] Goeree, J.K. and C.A. Holt (2001), Ten Little Treasures of Game Theory and Ten Intuitive Contradictions", American Economic Review 91, 1402-1422.

[16] Grimm, V. and F. Mengel (2008), An Experiment on Learning in a Multiple Games Environment, mimeo.

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

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

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

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

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

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

[23] Jehiel, P. and D. Samet (2005), Learning to play games in extensive form by valuation, Journal of Economic Theory 124, 129-148.

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

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

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

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

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

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

[30] Porta, J.M. (1999), " Reinforcement Based Learning with Automatic Categorization" Technical Report Instituto de Robótica e Informática Industrial, IRI-DT-9902.

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

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

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

[34] 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.

[35] 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.

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

[37] Steiner, J. and C. Stewart (2008), Contagion through Learning, Theoretical Economics, forthcoming.

[38] Van Damme, E., R. Selten and E. Winter (1990), Alternating Bid Bargaining with a Smallest Money Unit, Games and Economic Behavior 2, 188-201.

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

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

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

A Appendix: Some Definitions

Definition: Asymptotically Stable Point Let $\dot{x} = F(x)$ be a dynamical system in a set $\mathbf{X} \subset \mathbb{R}^c, c \in \mathbb{N}$. A restpoint x^* (i.e. a point s.t. $F(x^*) = 0$) is said to be asymptotically stable if the following two conditions hold.

(i) Given any neighborhood $\mathcal{N}_1$ of x^*, there exists some other neighborhood of it, $\mathcal{N}_2$, such that, for any path (or solution) $x(\cdot)$, if $x(0) \in \mathcal{N}_2$ then $x(t) \in \mathcal{N}_1$ for all t larger 0.

(ii) There exists a neighborhood $\mathcal{N}_3$ of x^* such that, for any path $x(\cdot)$, if $x(0) \in \mathcal{N}_3$, then $\lim_{t\to\infty} x(t) = x^*$.

Definition (Lyapunov Stability) A point is called Lyapunov stable if condition (i) from the previous definition is satisfied.

Definition (Unstable Restpoint) A restpoint (i.e. a point s.t. $F(x^*) = 0$) is called unstable if condition (i) from the previous definition is *not* satisfied.

B Appendix: Stochastic Approximation - Intuition and Proofs from Section 2

B.1 Mean Motion

It is intuitively clear that the mean motion of action choice frequency p_{mk}^{it} will depend on how much action a_m is reinforced in analogy class g_k compared to other actions. Denoting expectations by $\langle .. \rangle$, we can state the following Lemma.

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

$$\left\langle p_{mk}^{i(t+1)} - p_{mk}^{it} \right\rangle = \frac{1}{\beta_k^{it}} [p_{mk}^{it} r_k^{it} S_{mk}^{it}(x^t) + \varepsilon_0(1 - M p_{mk}^{it})] + O\left((\frac{1}{\beta_k^{it}})^2 \right). \tag{14}$$

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,

$$p_{mk}^{1(t+1)} - p_{mk}^{1t} = \begin{cases} \frac{\beta_{mk}^{1t} + \pi^1(a^t,\gamma^t) + \varepsilon_0}{\sum_{a_h \in A_1} \beta_{hk}^{1t} + \pi^1(a^t,\gamma^t) + M\varepsilon_0} - \frac{\beta_{mk}^{1t}}{\sum_{a_h \in A_1} \beta_{hk}^{1t}} & \text{if } g_k, a_m \in w^{it} \\ \frac{\beta_{mk}^{1t} + \varepsilon_0}{\sum_{a_h \in A_1} \beta_{hk}^{1t} + \pi^1(a^t,\gamma^t) + M\varepsilon_0} - \frac{\beta_{mk}^{1t}}{\sum_{a_h \in A_1} \beta_{hk}^{1t}} & \begin{array}{c}\text{if } g_k \in w^{it} \\ a_m \notin w^{it}\end{array} \\ \frac{\beta_{mk}^{1t} + \varepsilon_0}{\sum_{a_h \in A_1} \beta_{hk}^{1t} + M\varepsilon_0} - \frac{\beta_{mk}^{1t}}{\sum_{a_h \in A_1} \beta_{hk}^{1t}} & \text{if } g_k \notin w^{it} \end{cases} \tag{15}$$

or equivalently

$$p_{mk}^{1(t+1)} - p_{mk}^{1t} = \begin{cases} \frac{(1-p_{mk}^{1t})\pi^1(a^t,\gamma^t) + \varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h \in A_1} \beta_{hk}^{1t} + \pi^1(a^t,\gamma^t) + M\varepsilon_0} & \text{if } g_k, a_m \in w^{it} \\ \frac{-p_{mk}^{1t}\pi^1(a^t,\gamma^t) + \varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h \in A_1} \beta_{hk}^{1t} + \pi^1(a^t,\gamma^t) + M\varepsilon_0} & \begin{array}{c}\text{if } g_k \in w^{it} \\ a_m \notin w^{it}\end{array} \\ \frac{\varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h \in A_1} \beta_{hk}^{1t} + M\varepsilon_0} & \text{if } g_k \notin w^{it} \end{cases}. \tag{16}$$

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. 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 $\delta_{hm} = 1$ if $h = m$ and zero otherwise.

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:

$$\left\langle p_{mk}^{1(t+1)} - p_{mk}^{1t} \right\rangle = \sum_{\gamma_j \in g_k} f_j \sum_{G_l \in \mathcal{G}} q_l^{1t} \mathrm{I}_{kl}$$

$$\left[p_{mk}^{1t} \sum_{a_n \in A_2} \frac{(1-p_{mk}^{1t})\pi^1(a_m^1, a_n^2, \gamma_j)\sigma_{nj}^{2t} + \varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h \in A_1} \beta_{hk}^{1t} + \pi^1(a_m^1, a_n^2, \gamma_j) + M\varepsilon_0} \right.$$

$$\left. + \sum_{a_\eta \neq a_m \in A_1} p_{\eta k}^{1t} \sum_{a_n \in A_2} \frac{-p_{mk}^{1t}\pi^1(a_\eta^1, a_n^2, \gamma_j)\sigma_{nj}^{2t} + \varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h \in A_1} \beta_{hk}^{1t} + \pi^1(a_h^1, a_n^2, \gamma_j) + M\varepsilon_0} \right]$$

$$+ \left(1 - \sum_{\gamma_j \in g_k} f_j \sum_{G_l \in \mathcal{G}} q_l^{1t} \mathrm{I}_{kl} \right) \frac{\varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h \in A_1} \beta_{hk}^{1t} + M\varepsilon_0} \tag{17}$$

Denote $\beta_k^{1t} = \sum_{a_h \in A_1} \beta_{hk}^{1t}$ and remember that $r_k^{it} := \sum_{G_l \in \mathcal{G}, \gamma_j \in \Gamma} q_l^{it} f_j \mathrm{I}_{kl} \mathrm{I}_{jk}$ - where $\mathrm{I}_{kl} = 1$ if $g_k \in G_l$ and zero otherwise. $\sum_{G_l \in \mathcal{G}} q_l^{it} \mathrm{I}_{kl}$ is the frequency with which a partition containing g_k is used (in population i) and $\sum_{\gamma_j \in \Gamma} f_j \mathrm{I}_{jk}$ the (independent) probability that a game contained in g_k is played. Then this can be rewritten concisely as follows,

$$\left\langle p_{mk}^{1(t+1)} - p_{mk}^{1t} \right\rangle = \frac{1}{\beta_k^{1t}} [p_{mk}^{1t} r_k^{1t} S_{mk}^{1t}(\cdot) + \varepsilon_0(1-Mp_{mk}^{1t})] + O\left((\frac{1}{\beta_k^{1t}})^2 \right). \tag{18}$$

To see that the difference between the first term in (18) and expression (17) is indeed of order $(\frac{1}{\beta_k^{1t}})^2$ note that,

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

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

$$- \frac{\varepsilon_0(1-Mp_{mk}^{1t})(1+\frac{M\varepsilon_0}{\beta_k^{1t}})^{-1}}{\beta_k^{1t}}$$

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

■

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*

$$\left\langle q_l^{i(t+1)} - q_l^{it} \right\rangle = \frac{1}{\alpha^{it}}[q_l^{it} S_l^{it}(x^t) + \varepsilon_1(1 - Lq_l^{it})] + O\left((\frac{1}{\alpha^{it}})^2 \right). \tag{19}$$

Proof. The changes in partition choice probabilities are given by

$$q_l^{1(t+1)} - q_l^{1t} = \begin{cases} \frac{(1-q_l^{1t})\left(\pi^1(a^t,\gamma^t)-\Xi(Z_l)\right)+\varepsilon_1(1-Lq_l^{1t})}{\sum_{G_h\in\mathcal{G}} \alpha_h^t+\pi^1(a^t,\gamma^t)+L\varepsilon_1} & \text{if } G_l \in w_1^t \\ \frac{-q_l^{1t}\left(\pi^1(a^t,\gamma^t)-\Xi(Z_h)\right)+\varepsilon_1(1-Lq_l^{1t})}{\sum_{G_h\in\mathcal{G}} \alpha_h^t+\pi^1(a^t,\gamma^t)+L\varepsilon_1} & \text{if } G_l \notin w_1^t \end{cases} \tag{20}$$

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 $\sum_{\gamma_j \in \Gamma} f_j \sum_{A_1 \times A_2} \sum_{G_h \neq G_l} q_h^t \left(\sum_{g_k \in G_h} p_{mk}^{1t} \mathrm{I}_{jk} \right) \sum_{a_n \in A_2} \sigma_{nj}^{2t}$. Multiplying delivers

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

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

$$\left\langle q_l^{1(t+1)} - q_l^{1t} \right\rangle = \frac{1}{\alpha^{1t}}[q_l^{it} S_l^{it}(x) + \varepsilon_1(1 - Lq_l^{it})] + O\left((\frac{1}{\alpha^t})^2 \right). \tag{21}$$

■

Proof of Proposition 1:

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

$$\begin{aligned} p_{mk}^{i(t+1)} &= p_{mk}^{it} + \frac{1}{\beta_k^{it}} \widetilde{Y}_{mk}^{it} \\ q_l^{i(t+1)} &= q_l^{it} + \frac{1}{\alpha^{it}} Y_l^{it} \end{aligned} \tag{22}$$

$\forall i = 1, 2, \forall a_m \in A_i, \forall g_k \in \mathcal{P}^+(\Gamma), \forall G_l \in \mathcal{G}$. 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 $\{v^{it}\}_t$ and $\{\widetilde{v}^{it}\}_t$ 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 $\widetilde{y}^i_{mk}(x^t) = p^{it}_{mk} r^{it}_k S^{it}_{mk}(\cdot) + \varepsilon_0(1 - Mp^{it}_{mk})$ and $y^i_l(x^t) = q^{it}_l S^{it}_l(\cdot) + \varepsilon_1(1 - Lq^{it}_l)$ are the mean motions derived before. Taking into account the normalization the unique step size of order t^{-1}, $\beta^{it}_k = \alpha^{it} = 1/(\mu + t\theta)$ can be substituted in (22). 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):** $\sum_t \frac{1}{\mu+t\theta}\left|v^{it}\right| < \infty$ 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 $\left(\dot{p}^i_{mk} = \widetilde{y}^i_{mk}(x)\right), \forall a_m \in A_i, g_k \in \mathcal{P}^+(\Gamma)$ and $\left(\dot{q}^i_l = y^i_l(x)\right), \forall G_l \in \mathcal{G}, i = 1, 2$ as standard results in stochastic approximation theory show.[35] ■

C Appendix: Proofs from Sections 3 and 4

In some of the proofs below we will use the fact that agents will choose a best response at any stable point x^* also in "off equilibrium" analogy classes, i.e. those that are visited with probability $O(\varepsilon_1)$. We are entitled to do so by Assumption 1. We now quickly explain why this is so and why we think Assumption 1 makes sense. Note that at any restpoint

$$\frac{-p^i_{mk} S^i_{mk}}{(1 - Mp^i_{mk})} = \frac{\varepsilon_0}{r^i_k}.$$

Now if analogy class g_k is "off the equilibrium path", i.e. visited with probability $O(\varepsilon_1)$, then the RHS can be rewritten $\frac{\varepsilon_0}{O(\varepsilon_1)}$. This expression - if ε_0 and ε_1 tend to zero at the same rate - will be positive, finite and bound away from zero. But then, if a_m is not a best response (and thus $S^i_{mk} \leq 0$) we need $p^i_{mk} \to 0$. On the other hand if a_m is a best response (and thus $S^i_{mk} \geq 0$), we need p^i_{mk} to be bound away from zero. Finally, if ε_1 tends to zero at a faster rate than ε_0 the RHS diverges and all actions will be used with uniform probability in "off equilibrium" analogy classes. Then the predictions can be different from Nash equilibrium (even for arbitrarily small reasoning costs), but it seems a rather trivial way to obtain such predictions.

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

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 immediate. 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 opponent's play. But then $\exists \overline{\xi} < \phi$ such that $\forall \xi < \overline{\xi} : \Pi_l^i(\widehat{x}) - \Pi_h^i(\widehat{x}) = \phi - (\Xi(Z_l) - \Xi(Z_h)) > 0$. Thus $\widehat{x}$ cannot be a stable restpoint. ■

Proof of Claim 1:

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 all games $\gamma \in \Gamma_2$. Consider 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 payoff difference between partition G_3 and all other partitions on average at $\widehat{x} : S_3^{it}(\widehat{x}) = \Xi(\widehat{x}) - \Xi(2) - 5\varepsilon_1 > 0$, as the coarse partition must have probability zero at $\widehat{x}$ (and thus $\Xi(\widehat{x}) > \Xi(2)$). If $f_2 > f_3$ the same is true for G_2 and if $f_3 > \max\{f_1, f_2\}$ it is true for either G_2 or G_3. Consequently $q_1 > O(\varepsilon_0)$ cannot be a stable restpoint of (5)-(6). 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 $p = 1/2$ in $\{\gamma_2, \gamma_3\}$ and player 2 in analogy class $\{\gamma_{1,}\gamma_3\}$. It then follows immediately by arguments analogous to those above that $G_2 \notin \operatorname{supp} q^{1*}$ and $G_3 \notin \operatorname{supp} q^{2*}$. 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,

$$\begin{pmatrix} 2(f_1 + f_3) + f_2 & 2f_2 + f_1 + f_3 \\ 2f_2 + f_1 + f_3 & 2(f_1 + f_3) + f_2 \end{pmatrix} \text{ for player 1} \tag{23}$$

and

$$\begin{pmatrix} 2f_1+f_2+f_3 & 2(f_2+f_3)+f_1 \\ 2(f_2+f_3)+f_1 & 2f_1+f_2+f_3 \end{pmatrix} \text{ for player 2.} \tag{24}$$

Given the assumption that $f_j < 1/2$ for $j = 1, 2$ - (23) and (24) 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 (23) - (24). The Jacobian matrix associated with the linearization of the perturbed dynamics at the equilibrium $(p^1, p^2) = (\frac{1}{2}, \frac{1}{2})$ is given by

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

with spectrum $\left\{\frac{1}{2}(-4\varepsilon_0 \pm \sqrt{(f_1+f_3-f_2)(f_1-f_2-f_3)+16\varepsilon_0^2})\right\}$.Given our assumptions on f_j the term under the square root is negative and thus both eigenvalues have strictly negative real parts.[36] 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 $\operatorname{sign}[O(\varepsilon_0)] \gtreqless 0 \Leftrightarrow p^i_{mk} \lesseqgtr \frac{1}{2}$ we know that x^* is asymptotically stable. ■

Proof of Claim 2:

Proof. (i) First we show that no stable point can induce the equilibrium (A, a) in γ_1. Note that whenever $f_1 > -\frac{17}{6} + \frac{1}{6}\sqrt{409}$, action A is a best response of the row player in the "average game" to player 2's equilibrium behavior at any such point (as by Proposition 2, a Nash equilibrium has to be induced in γ_2). But then at any point x where (A, a) is induced in γ_1: $S_C(x) = \Xi(2) - \Xi(1) > 0$ destabilizing any such point. Furthermore no point where either player holds the coarse partition can induce a Nash equilibrium in both games and thus (by Proposition 2) can't be stable. Finally if ξ is high ($\xi > \overline{\xi}(\Gamma)$), both players will use the coarse partition, but then whenever $f_1 < \frac{2}{3}$, action a is not a best response to A for player 2.

(ii) Now we show that the point x^* is asymptotically stable. First note that as $\operatorname{supp} E^{Nash}(\gamma_1) \cap \operatorname{supp} E^{Nash}(\gamma_2) = \varnothing$, using the coarse partition will induce a strict payoff loss. But then for ξ small enough, $S_C(x^*) < 0, \forall x \in \mathcal{N}_{x^*}$. On the other hand asymptotic stability of $\left(\frac{1}{2}A \oplus \frac{1}{2}B, \frac{1}{2}a \oplus \frac{1}{2}b\right)$ in γ_2 follows from arguments analogous to those developed in part (ii) of Claim 1 (Note that the third action C (c) is strictly dominated for both players in γ_2). Stability of (C, c) in γ_1 follows from standard arguments. See e.g. Proposition 5.11 in Weibull (1995). ■

Proof of Proposition 3:

Proof. (i) As $\operatorname{card}\Gamma = 1$ 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 Weibull (1995).

[36] 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.

(ii) Let the games have payoff matrices given by

$$\gamma_1 : \begin{array}{|c|c|c|c|} \hline & H & L & .. \\ \hline H & a_1, a_1 & a_2, a_3 & .. \\ \hline L & a_3, a_2 & a_4, a_4 & .. \\ \hline \vdots & \vdots & \vdots & \\ \hline \end{array}, \gamma_2 : \begin{array}{|c|c|c|c|} \hline & H & L & .. \\ \hline H & b_1, c_1 & b_2, c_3 & .. \\ \hline L & b_3, c_2 & b_4, c_4 & .. \\ \hline \vdots & \vdots & \vdots & \\ \hline \end{array}, \tag{25}$$

where (H, H) is a strict Nash equilibrium. As we want γ_2 to have a unique equilibrium in mixed strategies, that is stable to learning in a single game we assume wlg that $b_1 > b_3, b_4 > b_2, c_1 < c_3$, that $c_4 < c_2$ and that no other equilibria exist. (As will become clear below, the arguments in the proof extend directly to mixed equilibria with more than two actions in their support). (See part (ii) of the proof of Claim 2). Assume also $f_1/f_2 \in (\underline{f}(\Gamma), \overline{f}(\Gamma))$ where$\underline{f}(\Gamma)$ is such that strategy H is a best response in the average game to $\left(f_1 + f_2(\frac{c_4 - c_3}{c_1 + c_4 - (c_2 + c_3)})\right) H \oplus (1 - (\cdot))L$ for player 1 (row player). 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 $p^i_{H2} < \sigma^i_{H2}$, where σ^i_{H2} is the equilibrium frequency of action H in γ_2. But then for any player either $p^i_{HC} = 1$ is not a best response to the phenotypic play of player $-i$ or a strictly higher payoff is associated with the coarse partition at such an equilibrium. 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, $\left(\partial \dot{q}_C / \partial q_C\right) = \Xi(2) - \Xi(1) + O(\varepsilon_0) > 0$. The strict Nash equilibrium $\left(\sigma^1_{H1}, \sigma^2_{H1}\right) = (1, 1)$ cannot be induced at any stable restpoint. ■

Proof of Proposition 4:

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 (25). 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 $\lim_{\varepsilon_0 \to 0} \lambda_i(\varepsilon_0) > 0$.

(ii) Let the 2×2 game γ_2 be again the game described in (25). The mixed equilibrium of the average game is given by $\frac{f_1(a_4 - a_3) + f_2(b_4 - b_3)}{f_1(a_1 + a_4 - (a_2 + a_3)) + f_2(b_1 + b_4 - (b_2 + b_3))} = \frac{a_4 - a_3}{a_1 + a_4 - (a_2 + a_3)}$ given our assumptions.Consider the restpoint where both players hold the coarse partition and choose $p^i_{HC} = \widehat{\sigma}^1_{H1}$ 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 5:

Proof. (i) As $\text{card}\,\Gamma = 1$ 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 $\pi^i(a_d^i, x^{-i}, \gamma_1) - \pi^i(a_w^i, x^{-i}, \gamma_1) > 0, \forall x^{-i} \in \text{int}\,\mathbf{X}_{-i}$. Consider a restpoint $\widehat{x}$ that induces a_w^i. As $\widehat{x}$ is interior there exists a neighborhood $\mathcal{N}_{\widehat{x}}$ of $\widehat{x}$ s.t. $\pi^i(a_d^i, x^{-i}, \gamma_1) - \pi^i(x, \gamma_1) + O(\varepsilon_0) > 0, \forall x \in \mathcal{N}_{\widehat{x}} \cap \text{int}\,\mathbf{X}$ and consequently $\widehat{x}$ cannot be a stable restpoint.

(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 $\forall \xi > 0$, $\sum_\Gamma \left(\pi^i(\widehat{x}, x^{-i}, \gamma_j) - \pi^i(x, \gamma_j)\right) - (\Xi(1) - \Xi(x)) > 0$, $\forall x \in \mathcal{N}'_{x^*} \cap \mathbf{X}$, $i = 1, 2$. 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 (26) that $\dot{Q}(x^*, x) < 0$. Thus $Q(x^*, x)$ is a strict Lyapunov function and x^* asymptotically stable. ■

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. ■

Proof of Claim 3:

Proof. First note that choosing $a_i = 180$ is the only best response to $a_{-i} = 180$ in game γ_{180} and choosing $a_i = 300$ is the unique best response to $a_{-i} = 300$ in analogy class $\{\gamma_0, \gamma_5\}$. Denote by x^* the point from Claim 3. Next note that our assumptions imply that the expected payoff difference between G^* and the coarsest partition G_C $\Pi^i(x^*) - \Pi_C^i(x^*) \geq \min\{120 f_0 + 115 f_5, 180 f_{180}\} - (\Xi(2) - \Xi(1)) > 0$. The difference between G^* and the finest partition G_F is $\Pi^i(x^*) - \Pi_F^i(x^*) = \Xi(3) - \Xi(2) - 4 f_5 > 0$ given our assumptions on the reasoning cost. It is immediate to verify that all other possible partitions of cardinality two lead to strictly worse expected payoffs than G^*. Remember that restpoints are hyperbolic. But then, as expected payoffs are continous functions of x_{-i}, there exists an open neighborhood of x^*, denoted by $\mathcal{N}_{x^*}$ s.t. $\sum_\Gamma \left(\pi^i(x^*, x^{-i}, \gamma_j) - \pi^i(x, \gamma_j)\right) - (\Xi(2) - \Xi(x)) > 0, \forall x \in \mathcal{N}_{x^*} \cap \operatorname{int} \mathbf{X}, i = 1, 2$ ■

Proof of Claim 4:

Proof. Note that action choices at x^* are best responses in the singleton analogy classes. Also $f_2 > \frac{f_1}{15}$ is a sufficient condition s.t. choosing "Non-Nash" is the only best response to f_1"Top" $\oplus$ f_2"$Bottom$" for column players in analogy class $\{\gamma_1, \gamma_2\}$. Next note that the expected payoff difference for row players (denote $i = 1$) between G^{1*} and the coarsest partition G_C^1 is $\Pi^1(x^*) - \Pi_C^1(x^*) \geq \min\{35f_1, 20f_2 + 330f_3\} - (\Xi(2) - \Xi(1)) > 0$. As $\operatorname{card} A_1 = 2$ no partition can have higher cardinality than 2 for the row player (see the Remark following Proposition 6). And as all choices are best responses in each game taken separately all other partitions of cardinality two will yield strictly lower payoffs at x^*. Similarly for the column players (denote $i = 2$) the payoff difference between partition $G^{2*} := \{\{\gamma_1, \gamma_2\}, \{\gamma_3\}\}$ and the coarsest partition G_C^2 is $\Pi^2(x^*) - \Pi_C^2(x^*) = 280f_1 + 10f_2 - (\Xi(2) - \Xi(1))$ whereas the difference between the finest partition G_F^{2*} and partition G^{2*} equals $\Pi_F^2(x^*) - \Pi^2(x^*) = 20f_1 - (\Xi(3) - \Xi(2))$. The latter is strictly positive (under our assumptions) whenever $\widetilde{\xi} < 20f_1$. It then follows by analogous arguments to those employed several times above that x^* is asymptotically stable. ■

Proof of Claim 5:

Proof. Consider the Rubinstein Bargaining game with discount factor $\delta = \delta_1 f_1$. This is the expected discount factor when the games are not distinguished (and game γ_1 occurs with frequency f_1). Call this game the "average game". Player 1 chooses $a^1 = \left\lfloor \frac{1}{1+\delta_1 f_1} \right\rfloor_{\frac{1}{M}}$ and player 2 chooses $b^2 = 1 - \left\lfloor \frac{1}{1+\delta_1 f_1} \right\rfloor_{\frac{1}{M}}$ in any Nash equilibrium of the average game in which no player uses a strategy that is weakly dominated by some other pure strategy.[37] As (because of the perturbation) all restpoints are interior this implies that, given any payoff linear selection dynamics, these strategies will be observed with asymptotic probability one in the average game. (See e.g. Proposition 5.8 in Weibull (1995)).

Now we will show that there exists an asymptotically stable point x^* of the dynamics (5)-(6) in which both players hold the coarse partition with asymptotic probability one and choose $a^{1*} = \left\lfloor \frac{1}{1+\delta_1 f_1} \right\rfloor_{\frac{1}{M}}$ and $b^{2*} = 1 - \left\lfloor \frac{1}{1+\delta_1 f_1} \right\rfloor_{\frac{1}{M}}$ 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

[37] The proof goes as follows. Assume that agents are at a NE which is not a SPNE. If $b^{2*} > b^{2SPNE} \Rightarrow (b^{2*} - \frac{1}{M})$ weakly dominates b^{2*}. (For $a^1 \geq a^{1*}$ both strategies yield the same payoff whereas if $a^1 = a^{1*} + \frac{1}{M}$ strict gains will be made). If $b^{2*} < b^{2SPNE} \Rightarrow (b^{2*} + \frac{1}{M})$ weakly dominates b^{2*}. (If $b^2 \leq (b^{2*} + \frac{1}{M})$ there is no payoff difference, whereas if $b^2 > (b^{2*} + \frac{1}{M})$ strict gains will be made whenever $b^{1*} \leq b^{1SPNE}$. But if not $(b^{1*} - \frac{1}{M})$ weakly dominates b^{1*}).

g_U. "Gross" gains (not taking into account the reasoning cost) by using a finer partition are of order ε_0. But then net gains are negative, $\forall \xi > 0$. 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 $\forall \xi > 0$,

$$\begin{aligned} & \sum_{\Gamma} \left(\pi^i(x^*, x^{-i}, \gamma_j) - \pi^i(x, \gamma_j)\right) - (\Xi(1) - \Xi(x)) \qquad (26) \\ = \; & O(\varepsilon_0) - (\Xi(1) - \Xi(x)) > 0, \end{aligned}$$

$\forall x \in \mathcal{N}_{x^*} \cap \operatorname{int} \mathbf{X}, i = 1, 2$. A strict Lyapunov function can be found as above. ■

D Appendix: Proofs from Section 5

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.[38] 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 (9) 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

$$q_l^{it} = \frac{\exp^{\varepsilon_1^{-1} \overline{\Pi}_l^{it}(\cdot)}}{\sum_{G_h \in \mathcal{G}} \exp^{\varepsilon_1^{-1} \overline{\Pi}_h^{it}(\cdot)}} =: B_l(\cdot) \qquad (27)$$

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

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

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 $\varepsilon_1^{-1} =$

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

κ, 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 (6).[39] Note though that the noise term ($\varepsilon_1 \chi(q)$) is still decreasing in q_l. The first-order conditions for problem (10) 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

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

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 (5) 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 (5) - (6).

(ii) Now consider the process where agents correlate action and partition choice employing choice rule (11). The first-order conditions for problem (11) 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 (27) $\Pi_l^{it}(x^t)$ is used instead of the historical payoffs $\overline{\Pi}_l^{it}(\cdot)$. The stochastic approximation under choice rule (11) will coincide up to a multiplicative constant with that of rule (9)-(10). ■

Proof of Proposition 8:

Proof. It follows immediately from the argument developed in the proof of Proposition 2 above that this proposition continues to hold. Furthermore whenever card $\Gamma = 1$ the process SFP1 with choice rules (9) and (10) coincides with the process SFP2 with choice rules (9) and (12). 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

[39] 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.

example demonstrates that part (ii) of Proposition 3 can fail. Let two games occurring with the same frequency be given by

$$\gamma_1 : \begin{array}{|c|c|c|} \hline & \text{H} & \text{L} \\ \hline \text{H} & 2,2 & 3,1 \\ \hline \text{L} & 2,1 & 4,4 \\ \hline \end{array}, \gamma_2 : \begin{array}{|c|c|c|} \hline & \text{H} & \text{L} \\ \hline \text{H} & 5,5 & 1,0 \\ \hline \text{L} & 0,0 & 0,0 \\ \hline \end{array}.$$

(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 (9) and (12) to choice rule (13) is shown in analogy to the proof of Proposition 7. ■

E Appendix: Additional Experimental Results

In this section we briefly describe the other results (in the section Static Games with Complete Information) from the paper by Goeree and Holt (2001) and indicate how they might be given an interpretation in terms of learning across games.

Matching Pennies Games

They examine three matching pennies games with the following payoff matrix

	Left	Right
Top	$x, 40$	$40, 80$
Bottom	$40, 80$	$80, 40$

If $x = 80$ play in the laboratory almost perfectly conforms to the Nash equilibrium prediction of choosing either actions with equal probability. If $x = 44$ a large majority of row players choose "Bottom" whereas column players tend to choose "Left". If $x = 320$ row players tend to choose "Top" whereas column players choose "Right". Think of a set of games where $x \in [0, 350] \cap \mathbb{N}$ and imagine that all players hold partitions of cardinality two. One half of the players classify all games with $x < \overline{x}$ in one analogy class and those with $x \geq \overline{x}$ in another analogy class. The other half of players acts in the same way relying on a different threshold value $\overline{\overline{x}}$, though. By choosing the threshold values and reasoning costs appropriately an asymptotically stable point can be constructed consistent with the experimental observation.

Minimum Effort Games

In the minimum effort game two players have to simultaneously choose a level of effort a_i from the set $A = [110, 170] \cap \mathbb{N}$. Payoffs are given by

$$\pi_i(a, c) = 6 + \min\{a_i, a_j\} - ca_i,$$

where $c \in [0, 1]$ is the marginal cost of providing one unit of effort. All points where both players provide the same level of effort are Nash equilibria. If $c = 1$, though, choosing $a = 110$ weakly dominates all other effort choices. If $c = 0$ choosing $a = 170$ is a weakly dominant strategy. Goeree and Holt (2001) find that if $c = 0.1$ a majority of participants choose $a = 170$. On the other hand if $c = 0.9$ most participants choose $a = 110$. Goeree and Holt (2001) claim that "Nash theory is silent" about these results. Clearly, though, if one takes into account the possibility that players might classify games in different analogy classes, Nash theory can say something about these results.

Coordination Games

They also let subjects interact in two versions of the following Coordination Game

	L	H	S
L	$90, 90$	$0, 0$	$x, 40$
H	$0, 0$	$180, 180$	$0, 40$

.

In one treatment $x = 0$ and (almost) all participans choose action H. If $x = 400$ only around 67% of all row players and 75% of all column players choose H. Unfortunately Goeree and Holt (2001) do not report what the remaining column players choose. Again one could imagine that players classify games with different values of x in different analogy classes, leading to an increased number of L choices among row players as x rises.

Dynamic Games with Complete Information

Bargaining: Also Goeree and Holt (2001) describe results of experimental bargaining games. The games are two stage games (the entire pie is gone after stage 2) with a discount factor of $\delta = 0.4$ and $\delta = 0.1$ respectively between stage 1 and stage 2. Initially the pie has size 5\$. Average offers are 2.83 and thus close to the SPNE in the game where $\delta = 0.4$ and 3.40 and thus quite far away in the game where $\delta = 0.1$. A possible explanation is that in games in real life the pie does not shrink as quickly as it does in the games in the experiment and that subjects use a (somewhat) coarse partition. Unfortunately Goeree and Holt (2001) only provide averages of the first offers here, so a more precise explanation cannot be given.

"A Trust Game": As an example of a dynamic game with complete information Goeree and Holt (2001) create a game to see whether one "should trust others to be rational". In the game player 1 moves first and can choose between a "save" action S and a risky action R. If she chooses S the game is over. If she chooses R it is player 2's time to choose. Player 2 has two actions available P (punish) and N (Nash). The payoffs are as follows:

	P	N
S	$80, 50$	$80, 50$
R	$20, x$	$90, 70$

(R, N) is a strict Nash equilibrium and a SPNE of the corresponding extensive form game whenever $x < 70$. Goeree and Holt (2001) find that while if $x = 10$ almost all row players choose action R, if $x = 68$ only roughly half of them choose R. Whenever they are actually asked to make the choice a large majority of column players choose N.[40] This could be explained by learning across games if players face a set of games where $x \in \{0, 20, 68, 88\}$ where some players group games with high and low x in the same analogy class.

[40] To be precise if $x = 10$ 84% of row players choose R and when asked 100% of column players choose N. If $x = 68$ 48% of row players choose R and when asked 75% of column players choose N.