Transitions In 2x2 Zero-Sum Games

Manuel Villarreal & Arturo Bouzas
Facultad de Psicología, UNAM

Abstract

Studies in strategic interaction have shown that, when participants play a single game repeatedly, their behavior tends to approach an equilibrium. Different learning models have been proposed to account for the behavior changes that lead to equilibrium. However, in these settings, learning models tend to generate similar predictions about the dynamics of a player's behavior, making it difficult to contrast them. The present experiment tries to avoid this problem by introducing unsignaled changes in the payoff function of a game across a single session, in order to compare the predictions of three learning models to such abrupt changes.

Introduction

Learning models were introduced to behavioral game theory to model the behavior of participants that face the same game repeatedly. The main objective of these models is to account for the dynamics of participants' behavior as they gather information about the situation they face.

Models make different assumptions about the kind of information that players use and the variables that change as a function of experience. Learning models can be divided into two families: Reinforcement and Bayesian. Reinforcement models assume that a player's disposition to choose an alternative changes as a function of the gains received in any given repetition of a game. In contrast, Bayesian models assume that players form beliefs in the form of probability distributions about the behavior of their opponents, and that they choose on the basis of those beliefs.

One of the problems that arises when contrasting learning models on the basis of a single repeated game is that reinforcement and Bayesian models tend to make similar predictions about participants' behavior. However, these models' predictions differ when the enviroment in which they are applied changes over time (*e.g.* Gallistel, *et. al.*, *2014*). The present experiment contrasts three learning models in a reapeated Generalized Matching Pennies Game, where unsignaled changes in the payoff function are introduced. The models are:

- Reinforcement Learning:
  - Cummulative Reinforcement (Erev & Roth, 1998; CR).
- Bayesian Learning:
  - Stochastic Fictitious Play (Fudenberg & Kreps, 1993; SFP).
- Self Tunning Experience Weighted Attraction model (Ho, CamererSt-EWA)

Method

Participants:

- 70 participants between 18 and 25 years were matched at random at the start of the experiment.

Games:

- 4 versions of a Generalized Matching Pennies game for 100 trials. At each trial participants had 2 strategies available (Heads, Tails).

	H	T
H	4	-2
T	-2	2

Trials: 1-100

	H	T
H	17	-1
T	-1	1

Trials: 101-200

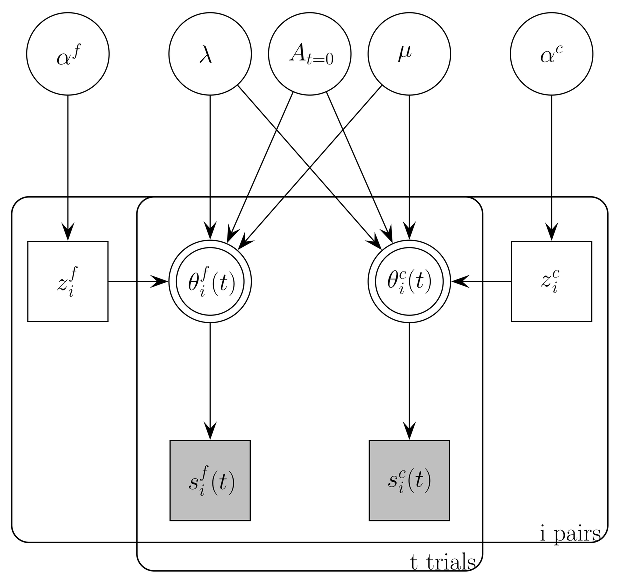
	H	T
H	6	-3
T	-3	3

Trials: 201-300

	H	T
H	8.5	-0.5
T	-0.5	0.5

Trials: 301-400

Models



$$\alpha^f\ ,\ \alpha^c \sim \text{Dirichlet}\big(\{1,1,1,1,1\}\big)$$

$$\lambda \sim \text{Gamma}\big(0.001, 0.001\big)$$

$$A_{t=0} \sim \text{Gamma}\big(0.001, 0.001\big)$$

$$\mu \sim \text{Gamma}\big(0.001, 0.001\big)$$

$$z_i^f\ ,\ z_i^c \sim \text{Categorica}\big(\alpha^{f,c}\big)$$

$$\theta_i^f(t), \theta_i^c(t) = \begin{cases} 0.5 & \text{si } z_i^{f,c} = 1 \\ CR(A_{t=0}) & \text{si } z_i^{f,c} = 2 \\ SFP(\lambda) & \text{si } z_i^{f,c} = 3 \\ St-EWA(\mu) & \text{si } z_i^{f,c} = 4 \\ Beta(1,1) & \text{otherwise} \end{cases}$$

$$s_i^f(t)\ ,\ s_i^c(t) \sim \text{Bernoulli}\big(\theta_i^{f,c}(t)\big)$$

CR:

$$\theta_i^j(t+1) = \frac{A_i^j(t)}{\sum_{k \in s_i} A_i^k(t)}$$

$$A_i^j(t) = A_i^j(t-1) + \begin{cases} R_i^j(s_i^j, s_{-i}) & \text{si } s_i^j = s_i(t) \\ 0 & \text{Otherwise} \end{cases}$$

SFP:

$$\theta_i^k(t+1) = \frac{e^{\bar{u}(s_i=k|\gamma_i(t))/\mu}}{\sum_{r \in S_i} e^{\bar{u}(s_i=r|\gamma_i(t))/\mu}}$$

$$\bar{u}(s_i = k|\gamma_i(t)) = \sum_{j \in S_{-i}} \gamma_i^j(t) * h_i(s_i = k, s_{-i} = j)$$

St-EWA:

$$\theta_i^j(t+1) = \frac{e^{A_i^j(t)/\mu}}{\sum_{k \in S_i} e^{A_i^k(t)/\mu}}$$

$$A_i^j(t) = \frac{\phi_i(t) A_i^j(t-1) N_i(t-1)}{N_i(t)} + \frac{[\delta_i(t) + (1-\delta_i(t)) I(s_i^j(t), s_i(t))] h(s_i^j, s_{-i}(t))}{N_i(t)}$$

Figure 1. Bayesian graphical model used to compare the learning models in the experiment. In order to assess the general quality of these, a random and a full model were included.

Results

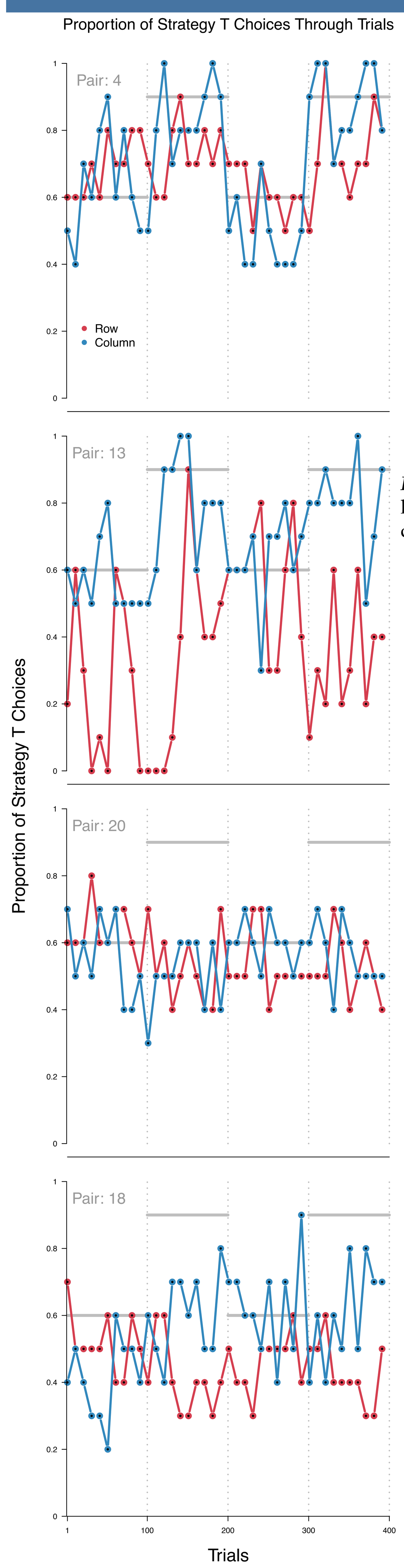


Figure 2. Four examples of participant's behavior. Proportion of choices of the strategy T by pair, the proportions were obtained for every 10 trials in the experiment.

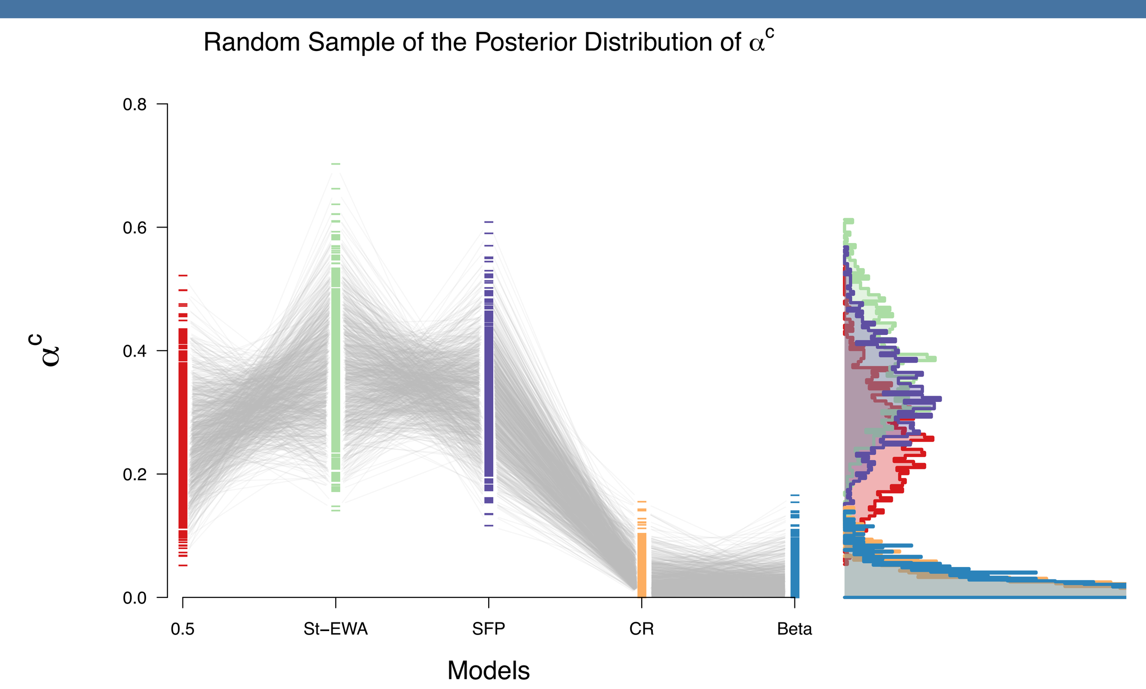


Figure 3. Random samples of the posterior distribution of the parameter α (probability of each learning model) for the coulumn players. The gray lines connect the five values of the components in α for a single iteration in a Markov Monte Carlo chain.

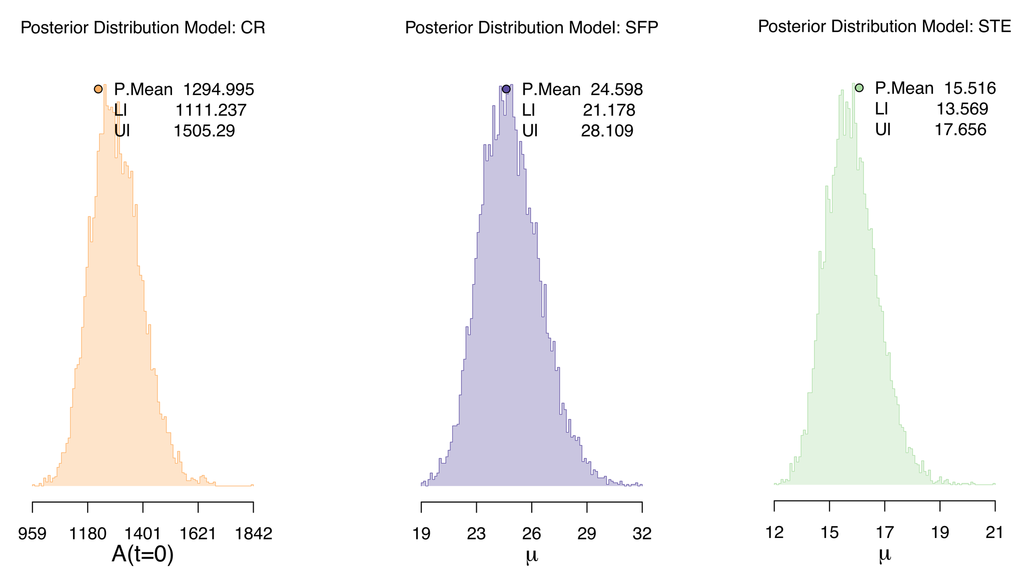


Figure 4. Histogram, mean and HPD interval for the posterior samples of the initial attractions in the CR model, anf the temperature in the SFP and St-EWA models.

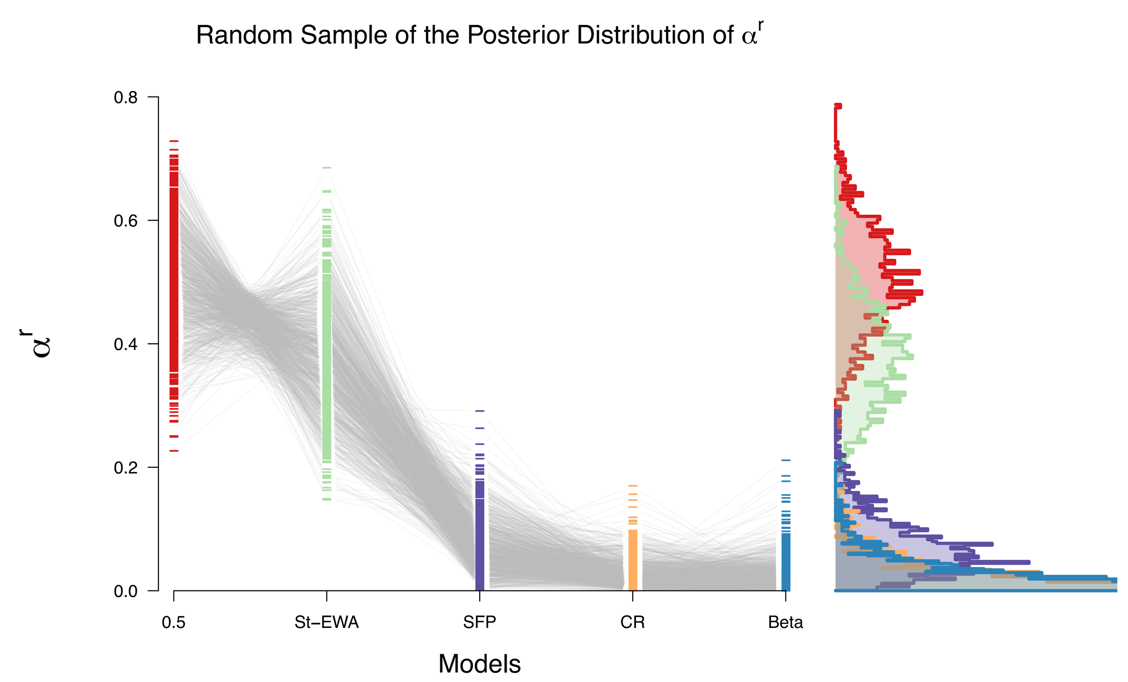


Figure 5. Random samples of the posterior distribution of the parameter α (probability of each learning model) for the row players. The gray lines connect the five values of the components in α for a single iteration in a Markov Monte Carlo chain.

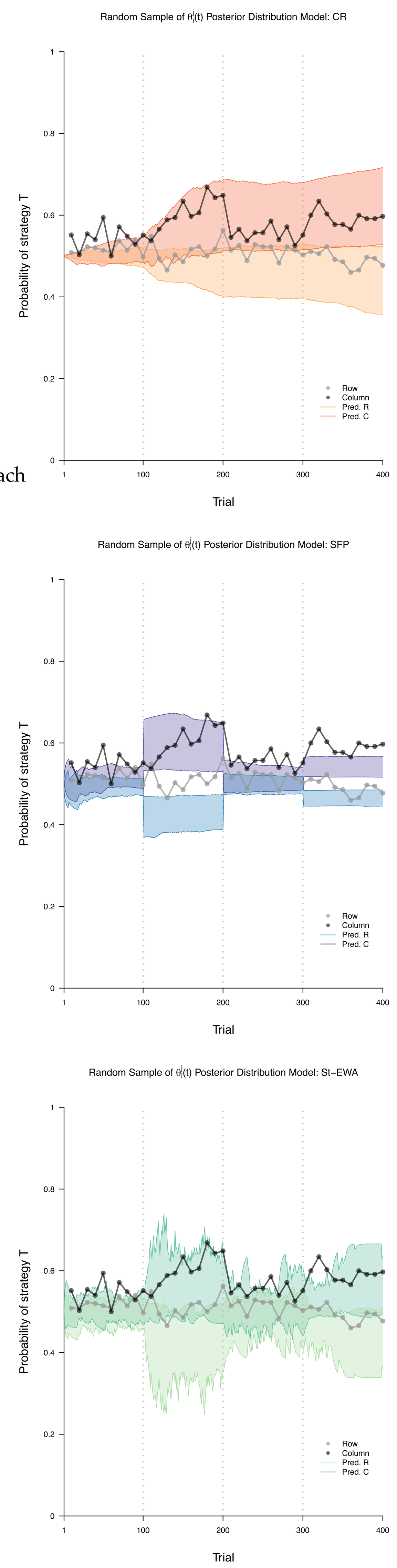


Figure 6. Random sample of the probability of choosing strategy T by row and column players for the three learning models.

Discussion

In the present experiment we found that, even though games were chosen to elicit the same kind of behavior in both players, the behavior of the participants differed. In general, the behaviour of the row players did not change with the changes in the payoff function, while the behavior of the column players did.

The main objective of the experiment was to contrast the prediction of three learning models when participants are in an environment (game) that changes over time.

Using a bayesian hierarchical model allowed to contrast the prediction of these three learning models with other alternatives, a random model and a full model; while taking into account the complexity of each one.

In this respect, the results suggest that a model that mixes the assumptions made by both families of models (model St-EWA) is more probable than the simplest forms of learning from both families.

Results also show that the most probable model differs from row to column players. The behavior of the row player is best accounted for by a model that predicts players will choose heads or tails with equal probabilities at every trial, while the behavior of the column player is better explained by models that assume choices are related to the opponent's behavior.

The predictions of the learning models differ in terms of the speed of change in participant's behavior, the magnitude of such a change, and its direction. Taking these three characteristics into account, the model that better describes the data is St-EWA.

Contact Information

Manuel Villarreal
jesus.muv@gmail.com
https://sites.google.com/site/adaptabilidad25

References

- Erev, I., y Roth, A. (1998). *Predicting How People Play Games: Reinforce- ment Learning in Experimental Games with Unique, Mixed Strategy Equilibria*. American Economic Review, 88(4), 848-881.
- Fudenberg, D., y Kreps, D. M. (1993). *Learning mixed equilibria*. Games and Econominc Behavior, 5, 320-367.
- Gallistel, C., Krishan, M., Liu, Y., Miller, R., y Latham, P. (2014). *The Perception of Probability*. Psychological Review, (121), 96-123.
- Ho, T., Camerer, C., y Chong, J. K. (2007). *Self-Tuning Experience Weighted Attraction Learning in Games*. Journal of Economic Theory, 133(1), 177-198.