Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

Reactive Versus Anticipative Decision Making in a Novel Gift-Giving Game

Elias Fernández Domingos,[1,2,3] **Juan Carlos Burguillo,**[3] and **Tom Lenaerts**[1,2]
[1] AI lab, Computer Science Department, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
[2] MLG, Département d'Informatique, Université Libre de Bruxelles, Boulevard du Triomphe CP212, 1050 Brussels, Belgium
[3] Department of Telematic Engineering. University of Vigo. 36310-Vigo, Spain
Corresponding: eliferna@vub.ac.be and tlenaert@ulb.ac.be

Abstract

Evolutionary game theory focuses on the fitness differences between simple discrete or probabilistic strategies to explain the evolution of particular decision-making behavior within strategic situations. Although this approach has provided substantial insights into the presence of fairness or generosity in gift-giving games, it does not fully resolve the question of which cognitive mechanisms are required to produce the choices observed in experiments. One such mechanism that humans have acquired, is the capacity to anticipate. Prior work showed that forward-looking behavior, using a recurrent neural network to model the cognitive mechanism, are essential to produce the actions of human participants in behavioral experiments. In this paper, we evaluate whether this conclusion extends also to gift-giving games, more concretely, to a game that combines the dictator game with a partner selection process. The recurrent neural network model used here for dictators, allows them to reason about a best response to past actions of the receivers (reactive model) or to decide which action will lead to a more successful outcome in the future (anticipatory model). We show for both models the decision dynamics while training, as well as the average behavior. We find that the anticipatory model is the only one capable of accounting for changes in the context of the game, a behavior also observed in experiments, expanding previous conclusions to this more sophisticated game.

Introduction

Many situations related to trust and fairness are modeled through gift-giving games, of which the Ultimatum and Dictator games are two examples (Forsythe et al. 1994; Cooper and Kagel 2009; Bardsley 2008; Nowak, Page, and Sigmund 2000; Fehr and Schmidt 1999; Kirchsteiger 1994). The Dictator Game (DG) takes place between pairs of individuals, where one is assigned the role of dictator and the other receiver. At the start of the game the dictator is assigned an endowment and she is requested to give an amount between 0 and the entire endowment to the receiver. The receiver has no choice in the matter: she just receives whatever the dictator gives. Game Theory predicts, under the assumption of rational and selfish behavior, that the dictator should keep everything for herself, which is referred to as the sub-game perfect Nash equilibrium of the game. Experiments have logically revealed that humans are not hyper-rational or fully selfish: 60% of the participants decide to share a positive amount of money with a mean transfer of approximately 28% of the endowment (Camerer 2003; Engel 2011). The Ultimatum Game (UG) differs from the DG in that the receiver obtains the power to refuse an offer, with the consequence that they both get zero payoff. This new bargaining power results into a significant increase in the amount offered, reaching almost half the endowment on average (Oosterbeek, Sloof, and Van De Kuilen 2004). Recently, a variation on these two games was proposed, i.e., the Anticipation Game (AG) (Zisis et al. 2015), to study the impact of group formation on generosity. The AG differs from the UG in the sense that the decision by the receiver is made prior to the offer of the dictator: The receiver is first informed about the reputation of the dictator, i.e., the amounts he or she gave in previous interactions. Using this information she then has to decide whether to play the DG with that particular player. If she refuses they both get zero payoff, if she agrees the payoff is defined by the outcome of the DG. Once the game is played the reputation of the dictator is updated by including the last donation that she made. The game has an equivalent sub-game perfect equilibrium as the UG and experiments show that also in that case, dictators give an amount close to an equal split of the endowment (Zisis et al. 2015).

To explain the origin of the behavior observed in the AG experiments, a stochastic evolutionary dynamics model was used similar to the one in (Rand et al. 2013). This model easily reproduces the average donations observed in different AG experiments using simple discrete strategies, yet does not capture closely the individual behaviors used by the participants (Fowler and Christakis 2013). To overcome this problem, the model had to incorporate in the dynamics that the participants have the capacity to anticipate (Zisis et al. 2015), i.e., they take into account that the amount they give will have an effect on being accepted in future interactions. This feature was implemented by simply assuming that the success of an individual is defined by the donation made now, the likelihood of being accepted in the next round and the payoff obtained for the same donation in the next round. Although this adaptation leads to a closer fit with the experimental results, it does not explain which cognitive mechanisms are necessary or which features they should have. Ad-

ditionally, in order to transfer the insights from these experiments to autonomous agent-based systems, it is important to make these mechanisms explicit, so that we can understand how the agents need to be implemented to display anticipatory behavior.

Prior works on anticipation in AI have discussed different approaches to model individual anticipatory behavior (Pezzulo et al. 2008). Some of these approaches use recurrent neural networks (RNN) (Elman 1990; Funahashi and Nakamura 1993) to determine the action that may lead to the best future outcome in the prisoners dilemma(Lalev and Grinberg 2006). In this paper we examine whether the results in that work expand to the AG, which is more complex in the sense that more actions are possible for both the dictator and receiver. We show that while the reactive model is capable of identifying optimal actions for simple scenarios (e.g., all the receivers have the same fixed strategy), it's performance decreases considerably when faced against more realistic situations. In contrast, the forward-looking model is able to respond to changes in the context of the game, which leads it to a more successful outcome. Furthermore, the RNN is able to effectively represent the consequences of the dictator's actions for several time-steps.

The remainder of this paper is organized as follows. First we discuss the prior work in modeling anticipation on which the current paper is based. Then, in the Methods section, we explain the reactive and anticipatory models used. Afterwards, we present the results obtained from simulations and relate them to the experimental results obtained in (Zisis et al. 2015). Finally, we draw some conclusions about the results and explain the next steps we will take in this line of research.

Related Work

Learning theory applied to social dilemmas focuses often on the Prisoners Dilemma (PD) and on reactive models to explain the observed behavior (Fudenberg and Levine 1998; Roth and Erev 1995; Börgers and Sarin 1997; Macy 1991; Macy and Flache 2002; Flache and Macy 2002; Masuda and Nakamura 2011; Ezaki et al. 2016). Reactive systems are those that respond or react to immediate environmental and internal needs. In consequence, they do not focus on future needs, but rather they have to wait for the conditions to occur first (Pezzulo et al. 2008). Macy and Flache, for instance, proposed a backward-looking (reactive) stochastic reinforcement learning model based on the Bush-Mosteller (BM) model, which adapts the probability of cooperation depending on the difference between an aspiration level and the received payoff. This aspiration level represents the threshold reward an agent is expecting to receive. Consequently, the *perceived* stimulus will be either negative or positive, if the payoff obtained at a certain instant falls below or above this threshold. The agent then reacts to these stimuli so that actions associated to positive sensed values are reinforced, in other words, the player will continue the action after gaining a relatively large payoff (which they call satisficing) and would switch otherwise. The authors identify a dynamic solution concept, and claim that it might lead adaptive agents out of the social trap of the PD and into stable mutual cooperation, a process they call stochastic collusion. However, they show that this process is highly sensitive to the dynamics of the aspiration. (Macy 1991; Macy and Flache 2002; Flache and Macy 2002).

Masuda and Nakamura argue that the BM player with a fixed aspiration level is essentially the same as the Pavlov strategy that only uses the information about the immediate past (Kraines and Kraines 1989; Nowak and Sigmund 1993). They also say that the BM model with adaptive aspiration level is not known to yield a large probability of mutual cooperation except in some limit cases. With this in mind, they analyse a slightly modified version of the model, in which the reinforcement signal has a stronger effect on the action selection. They show that for low to intermediate speeds of adaptation of the aspiration level, the modified BM player mutually cooperates with a large probability and is competitive in evolutionary dynamics, which may serve to explore the relationship between learning and evolution in social dilemmas (Masuda and Nakamura 2011). A more recent paper, applies this model to a multi-agent environment in which agents play the two-player PD against each (direct) neighbor in a square lattice. The agents behave similarly to conditional cooperators, a strategy identified in behavioral experiments, and so the authors claim that this aspiration learning model is able to give a proximate explanation to the conditional behavior of humans (Ezaki et al. 2016).

Despite the extensive use of backward-looking reactive models, some authors have argued that forward-looking or anticipatory learning models are more suitable to explain experimentally observed results. An anticipating system will be defined here as *a system containing a predictive model of itself and/or its environment, which allows it to change state at an instant in accord with model's predictions pertaining to a later instant* (Rosen 1985). A nice overview of anticipatory learning is provided in (Pezzulo et al. 2008). Of direct interest for the current work is (Taiji and Ikegami 1999): Taiji and Ikegame, presented "pure reductionist Bob", an anticipative agent that is able to generate a predictive model of the opponent, and "clever Alice", another agent that assumes the opponent is "pure reductionist Bob", thus, generates a predictive model of itself. Both models use a RNN (Elman 1990; Funahashi and Nakamura 1993; Bhatia and Goldman 2014). They show how the learning agent is able to select the optimal actions against players with a fixed strategy. Also, when playing among themselves, the anticipative agents reproduced complex dynamics that concluded with full defection. In (Lalev and Grinberg 2006), the authors compare the performance of an adapted version of the reactive model of Macy and Flache to an anticipatory model in the context of the two-player PD. Both models use the same RNN to generate predictions about the game. The backward-looking model uses the prediction of the payoff as the aspiration level, while the forward-looking model uses the network to predict several future outcomes and make a decision based on them. Lalev and Grinberg showed that both models produced no statistical difference to the experimental results in regards to the mean payoff. On the other hand, when comparing the mean levels of cooperation, only the anticipatory model resembled the experimental data, as

the levels of cooperation for the reactive model were higher. Moreover, when comparing the response of the models to changes introduced in the payoff matrices the reactive model gave a totally inadequate description of the experimental results, while the forward-looking model fitted them closely. A possible reason for this may be that the backward-looking mechanism only accounts for the previously received payoff and the expected payoff, thus not making use of the predictions about the change of context that the network was able to make. The authors concluded that the anticipatory model accounts better for human behavior. The RNN model of Lalev and Grinberg will be discussed in more detail in the next section as it is the foundation for the work we present here.

Finally, anticipative learning in combination with game theory was also applied to real world problems. For instance, (Slimani, El Farissi, and Achchab) show how it can be applied to study behavioral probabilities in a supply chain, where firms face the problem of predicting the market demand in order to start the production before orders are made. This is necessary, as manufactures cannot risk for the actual demand to occur to start reacting. However, the ones that posses a better view of the market are the retailers, who may choose to withhold the information, and making an error on the prediction here can cost too much. The author experimented with different architectures of multi-layer neural networks and were able to produce satisfactory results while forecasting demands.

Model and Methods

Modeling Receivers

To train the dictators, other agents that play the role of receiver need to be provided. Here the receivers are modelled using a fixed conditional strategy that specifies from which amount onwards she accepts a dictator. In the result section, these receivers will be used to train and test the dictators' performance using two scenario's: The first scenario considers only one receiver with a specific strategy. The second scenario assumes that there is a pool of receivers with different strategies that can be encountered with a particular probability, which corresponds to more realistic situations as observed in (Zisis et al. 2015). The decision to accept or not by the receiver depends on the availability of reputation information for the proposed dictator. In the final subsection of the result section, we will also examine how the dictator behaviour changes if that information is not always available.

Modeling Dictators

Although both dictators and receivers could be modeled using forward-looking learning, we will for now only focus on the modeling of the former. Both reactive and anticipative agents were developed, using the models in (Macy and Flache 2002) and (Lalev and Grinberg 2006) as a basis. We use a RNN to generate a predictive model of the player itself and of the environment. Then, each model makes use of this network differently to make decisions. We will explain this further in the next subsections.

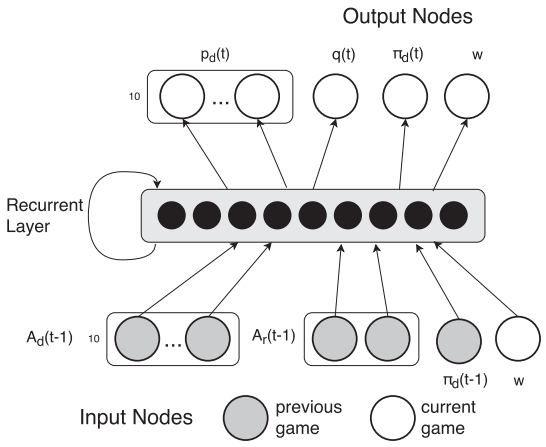

Figure 1: Schematic representation of the recurrent neural network used by the dictator's agent. Explanation is provided in the text.

The RNN used has 14 inputs and 13 outputs and 100 hidden nodes. We also tested other possible sizes for the hidden layer, however this was the one that showed a best response, without exceeding the constraints of our resources. The activation functions of the hidden layer and the output layer are tan-sigmoid and log-sigmoid respectively. All inputs to the network are normalized to the interval $[0, 1]$. Figure 1 represents schematically this network. The actions of the dictator and the receiver are encoded as orthonormal vectors on the inputs, so that each time only one input is active for each action. In the AG, the dictator has 10 possible actions (Zisis et al. 2015), ranging from 1 to 10, thus the input vector for the dictator's action is of size ten. Prior to the action of the dictator, the receiver has to decide to either accept or reject the interaction, so the input vector for her actions has size two. Yet note that when the action is *not accept* the input vector of the dictator is an array of zeros. This is because, the game does not proceed if the receiver does not accept the interaction. In Figure 1, the first ten inputs named $A_d(t-1)$ correspond to the dictator's action on the last game. $A_r(t-1)$ corresponds to the action of the receiver on round $t-1$ and $\pi_d(t-1)$ refers to the payoff received by the dictator on that round. Finally w represents the state of the game. If $w = 0$, the receivers have to choose their actions without having any information about the dictator, which according to the experiments, increases the probability of acceptance and also reduces the amount the dictators share (Zisis et al. 2015). When $w = 1$, receivers have information about the past 3 actions of the dictator when they were accepted by a receiver.

The outputs of the network are interpreted as probabilities. The first 10 outputs correspond to the expected value of each possible action of the dictator ($p(t)^{1-10}$). The 11th output refers to the probability of acceptance of the receiver, q_t. The 12th output is the predicted dictator's payoff for the round t, $\pi_d(t)$ The last output is again w, which corresponds to the output of the same name. This is intended to generate a representation of the state of the game within the network

as in (Lalev and Grinberg 2006).

The code used to implement the models presented can be found at https://github.com/Socrats/anticipation-matlab.

Training

The network is trained after every game with a sliding window that includes the last 10 epochs. The inputs to the network are always the last game outcome (dictator action, receiver action and payoff of the dictator) and the state at the present game w, and the targets are generated after the players finish the game: the target for the predicted probability of acceptance, $q(t)$, is exactly the action of the receiver. The payoff ($\pi_d(t)$) corresponds to the payoff of the dictator in this game for the Anticipatory model and to the average of the payoff through the past games for the reactive model. The target probability for each action of the dictator ($p(t)^{1-10}$) is calculated using one of the two methods presented in the following subsections.

Reactive model

The backward-looking reinforcement learning proposed in (Macy 1991) is a variant of the Bush-Mosteller model (Bush and Mosteller 1953). The model consists of a stochastic decision rule and a learning algorithm in which the consequences of the decision create positive and negative stimuli (rewards and punishments) that update the probability p that the decision will be repeated.

Originally, this model was defined for the PD game, which comprises of only two possible actions. We have adapted this model to our specifications: Given that our game has 10 different actions, the following extension of the model needs to be used:

$$s(t) = \frac{r(t) - Asp(t)}{sup[|max(r) - Asp(t)|, |min(r) - Asp(t)|]} \quad (1)$$

In Equation 1, $s(t)$ is the sensed stimulus, $r(t)$ is the payoff obtained and $Asp(t)$ is the aspiration level of the player at round t. The aspiration level indicates the threshold for which the player satisfices, which means that the player will increase the probability of repeating the previous action. Therefore, the dynamics of this threshold are also important to understand the outcome of the game and the strategies selected by the player. Here, we use the payoff predictions of the RNN defined in the previous subsection as a representation of the player's expectations and use it to update the aspiration level in the model, similarly to what is done in (Lalev and Grinberg 2006). The probabilities of each action are updated according to Equation 2 when the action to update is the last selected action ($A(t-1) = i$) and according to Equation 3 when it is not ($A(t-1) = j, j \neq i$).

$$p_i(t) = \begin{cases} p_i(t-1) + (1 - p_i(t-1))ls(t) & (s(t-1) \geq 0), \\ p_i(t-1) + p_i(t-1)ls(t) & (s(t-1) < 0) \end{cases} \quad (2)$$

$$p_i(t) = \frac{(1 - p_j(t))p_i(t-1)}{\sum_{n=1, n\neq j}^{N} p_n(t-1)} \quad (3)$$

N is the number of actions, $p_i(t)$ is the probability of action i at round t and l is the learning rate.

Anticipatory Model

Here we explain the anticipatory agent model, which uses the RNN to estimate the impact each action may have on the future. The decision-making process then decides what is the best strategy to take on the present, given those future outcomes.

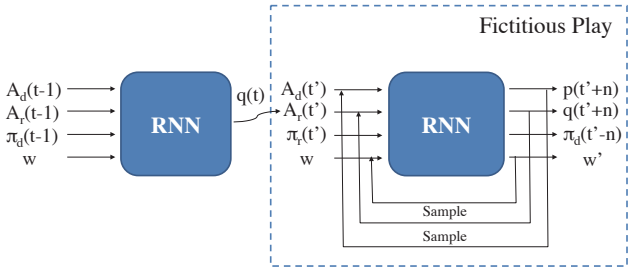

Figure 2: Schema of the anticipation process. Explanation is provided in the text.

In order to assess the outcome of a potential action, the dictator generates a sequence of n elements using the RNN under the assumption that she uses that particular action now. The initial input for the RNN consist of this particular action, a prediction of the receiver's action and a payoff. The latter is calculated by $\pi_d(t') = 10 - A_d(t')$, where $A_d(t')$ is the dictator's action at time t', of the fictitious play (see Figure 1). The action of the receiver at time t' is obtained by feeding the past game information (actions of both players, payoff and state at time $t-1$), generating a prediction of the strategy of the opponent in form of a probability distribution (i.e., $q(t)$ in Figure 1). As we mentioned before, the reason the outputs of the network are interpreted as probabilities is related to the sigmoid function that activates the outputs and that squashes them into the interval $[0, 1]$. It is important to notice that there is zero payoff when the receiver rejects an interaction. In the following $(n-1)$ steps of this fictitious play, the actions of the dictator are taken by choosing the action with maximum expected value, and the actions of the receiver are generated by sampling from the distribution of probabilities predicted by the RNN. At each step we also calculate the expected payoff with the utility function: $u(t') = \pi_d(t')q(t')$ This process is repeated until $t' + n$. The same process is executed for every action the dictator can take. Once the sequences and corresponding payoffs are obtained, the expected values, $Q_i(t)$, are calculated using Equation 4. In this equation, λ is a discount factor that defines the importance of the future predictions to the probability. Finally, the probabilities for each possible action of the dictator are calculated with Equation 5, wherein β represents the sensitivity towards the difference in payoff. The bigger β is, the more sensible to this difference in payoff. Figure 2 visualizes the entire process.

$$Q_i(t) = u_i(1) + \sum_{k=2}^{n} u_i(k)\lambda^{k-2} \quad (4)$$

$$p_i(t) = \frac{e^{Q_i(t)\beta}}{\sum_{j=1}^{N} e^{Q_j(t)\beta}} \quad (5)$$

Results

Both reactive and anticipatory models identify the required donation when the receiver strategy is fixed

As a first step, we study the performance of dictators with reactive (see Figure 3) and anticipatory (see Figure 4) models by playing against a receiver with a fixed strategy, which is in this case accepting more than 2. As a consequence, the maximum possible average payoff that the dictator can obtain is 7. In the Figures, we show the dynamics of the probability distribution over the actions for each round during the training process. We can see that in both models the probability distribution converges to the best choice (i.e., giving 3), however the reactive model takes longer to reach that equilibrium. In Figure 3 (embedded plot) we can see that the dictator's aspiration level is converging very slowly towards 7. It is important to notice that in both models we selected the best parameter settings, which is $l = 1$ for the reactive model, and $\lambda = 0.3$ and $\beta = 1/0.01 = 100$ for the anticipating model.

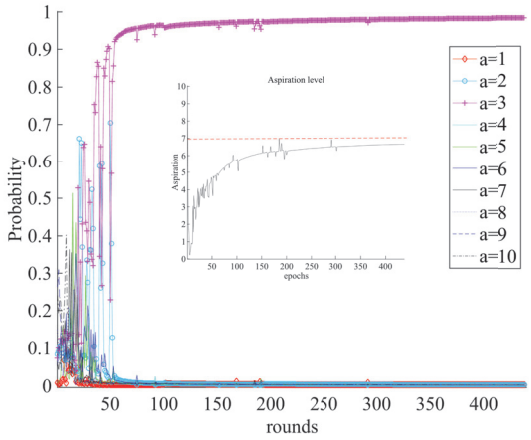

Figure 3: Variation of the probability of each action during training for the reactive model. The dictator plays only against receivers that accept anything above 2 and that will always accept the dictator if she was rejected by the receiver on the previous round. We set $l = 1$.

Playing against different receiver types leads also to similar results

Here, we study the behavior of the models in a more realistic scenario: the dictators play against receivers with different strategies drawn randomly (with a certain probability distribution) from a pool. Concretely, the receivers in the pool

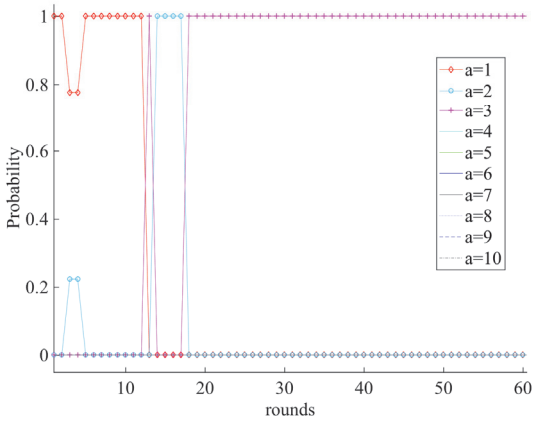

Figure 4: Variation of the probability of each action during training for the anticipatory model. The dictator plays against a pool of receivers that accepts anything above 2 and that will always accept the dictator if she was rejected by the receiver on the previous round. We set $\lambda = 0.3$ and $\beta = 100$.

follow one of 5 possible strategies, named strategy S_1 to S_5, which differ in the minimal donation they accept from the dictator. This acceptance threshold ranges from accepting everything to accepting anything above 4, respectively. We determined the results for two cases: the receivers are drawn with uniform probability from the pool and the receivers are drawn according to the following probability distribution: 0.2 for strategy S_1 , 0.5 for S_2 , 0.15 for S_3 , 0.1 for S_4 and 0.05 for S_5. We call this latter scenario *non uniform* for the sake of legibility. We used the same parameter settings for each model as in the previous subsection.

In this case, we first trained the reactive and anticipating dictators until they converged and, afterwards, they play 10 times for 30 rounds against the receivers from the pool. In Figure 5, we show the average payoff obtained over this 10 runs for the 2 scenarios we mentioned. The maximum average payoff that could be obtained for the scenario with uniform probability is 7 and for the other is 7.7. From the results, we can see that both dictators obtain an average payoff very close to 5 in both scenarios. This outcome is sensible as it is not possible for the dictators to predict the strategy of each receiver individually, even for the non uniform case. Therefore, the best is to try to maximize the acceptance, as not being accepted means receiving 0 payoff, and always donating an amount close to 5 is the only strategy that guarantees full acceptance.

Anticipatory model identifies changes in the game context

The experiments from (Zisis et al. 2015) show that the average acceptance increases when the receivers have partial or no information about the reputation of the dictator they are facing. Additionally, the dictators tend to share more of the endowment when they know receivers can see their

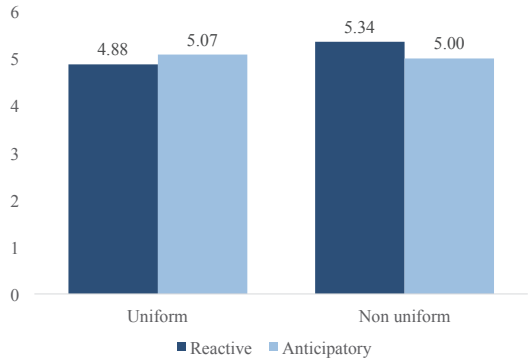

Figure 5: Average payoff obtained by the dictators while playing against a pool of receivers that have different strategies. The uniform case means that the receivers are drawn with uniform probability from the pool and the non uniform means that they are drawn following the probability distribution explained in the text. In the 2 cases and for both models, the average payoff is similar and is close to 5.

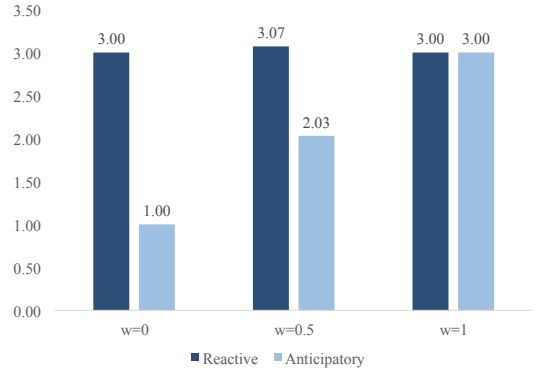

Figure 6: Average donation for each model for three different contexts: $w = 0$, $w = 0.5$ and $w = 1$. The reactive and anticipative dictators were first trained playing against a pool of receivers that accept anything above 0, 1 and 2 respectively for each context. The context changed randomly during the training every 5 rounds. The results were obtained by letting the trained dictators play for 30 rounds against the receivers on each context, and then averaging the results over 10 runs.

history of past actions. Now, we show the behavior of both the reactive and anticipatory models when facing a similar situation. We train the reactive an anticipating dictators in 3 different contexts: $w = 1$, when the receivers have all the information about the dictator and in this case they only accept shares above 2 units; $w = 0.5$, in the experiments this means the receivers have all the information only half of the time, we modeled this by making the receivers accept lower endowments (of above 1 unit); and $w = 0$, when the receivers have no information about the dictator, and in this case they accept any share from the her. In Figure 6 we show for each context the average amount the dictators share. We can see how the anticipatory model performs well and is able to detect the changes in context which allow her to make the optimum action almost each time. In contrast, the reactive model is not able to account for the change of context, despite using the same recurrent neural network as the anticipatory model and having the same inputs.

Conclusions

We have compared a reactive and an anticipatory model applied to a novel gift-giving game, i.e., the AG. We used a RNN to generate a predictive model of the player, the opponent and the context of the game. The reactive dictators use this RNN in a backward-looking reinforcement learning model based on aspirations. This aspiration was updated based on the predicted payoff generated by the RNN, as proposed in (Lalev and Grinberg 2006). The anticipative dictators, used the network to generate the distribution of probabilities for the available actions through fictitious play, so that actions that lead to better future rewards would have greater probability of being chosen.

The results we obtained have shown that, in general, the anticipatory model is able to learn the best action much faster than the reactive model. In addition, the later model was not able to perform well in more complex environments, where receivers play different strategies. More importantly, the anticipative dictator was able to identify the change in game context and play accordingly. In contrast, the reactive agent converged to a suboptimal state, that is, she chose as action the minimum amount required to be accepted in the most restrictive context. Therefore, the anticipatory model can describe better the experimental results in (Zisis et al. 2015). This is relevant as it gives a proximate confirmation for the results, where the levels of dictators' donations decrease when they know the receivers have no information about their reputation. Thus, anticipation, together with the presence of reputation, truly guides participants towards a more fair outcome.

Our results highlight the importance of anticipation for human cognition. However, the capacity to anticipate also comes with a cost. In order to plan or reason about the consequences of their actions through anticipation, individuals need to build useful model representations of the environment (Pezzulo et al. 2008). In more complex environments, individuals might be mislead by wrong predictions that can lead them into suboptimal or harmful states. Nevertheless, these drawbacks are minor in comparison to the potential advantages of anticipatory reasoning and its relation to the capacity of humans to accomplish simple and complex forms of social interaction.

Acknowledgments

EDF and TL are supported by the grant FRFC nr. 2.4614.12 from the Fondation de la Recherche Scientifique - FNRS and the grant nr. G.0391.13N provided by Fonds voor Wetenschappelijk Onderzoek - FWO. JCB is supported by the University of Vigo.

References

Bardsley, N. 2008. Dictator game giving: altruism or artefact? *Experimental Economics* 11(2):122–133.

Bhatia, S., and Goldman, R. 2014. A Recurrent Neural Network for Game Theoretic Decision Making. In *Proceedings of the 36th annual conference of the Cognitive Science Society*.

Börgers, T., and Sarin, R. 1997. Learning through reinforcement and replicator dynamics. *Journal of Economic Theory* 77(1):1–14.

Bush, R., and Mosteller, F. 1953. A stochastic model with applications to learning. *The Annals of Mathematical Statistics* 24(4):559–585.

Camerer, C. 2003. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Cooper, D., and Kagel, J. H. 2009. Other regarding preferences: a selective survey of experimental results. *Handbook of experimental economics* 2.

Elman, J. L. 1990. Finding structure in time. *Cognitive Science* 14(2):179–211.

Engel, C. 2011. Dictator games: A meta study. *Experimental Economics* 14(4):583–610.

Ezaki, T.; Horita, Y.; Takezawa, M.; and Masuda, N. 2016. Reinforcement learning explains conditional cooperation and its moody cousin. *PLoS Comput Biol* 12(7):e1005034.

Fehr, E., and Schmidt, K. M. 1999. A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics* 114(3):817–868.

Flache, A., and Macy, M. W. 2002. Stochastic collusion and the power law of learning a general reinforcement learning model of cooperation. *Journal of Conflict Resolution* 46(5):629–653.

Forsythe, R.; Horowitz, J. L.; Savin, N. E.; and Sefton, M. 1994. Fairness in simple bargaining experiments. *Games and Economic behavior* 6(3):347–369.

Fowler, J. H., and Christakis, N. A. 2013. A random world is a fair world. *Proc. Natl. Acad. Sci. USA* 110(7):2440?2441.

Fudenberg, D., and Levine, D. K. 1998. *The theory of learning in games*, volume 2. MIT press.

Funahashi, K., and Nakamura, Y. 1993. Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks* 6(6):801–806.

Kirchsteiger, G. 1994. The role of envy in ultimatum games. *Journal of Economic Behavior and Organization* 25(3):373–389.

Kraines, D., and Kraines, V. 1989. Pavlov and the prisoner's dilemma. *Theory and Decision* 26(1):47–79.

Lalev, E., and Grinberg, M. 2006. Backward vs. forward-oriented decision making in the iterated prisoner?s dilemma: A comparison between two connectionist models. In *Workshop on Anticipatory Behavior in Adaptive Learning Systems*, 345–364. Springer.

Macy, M. W., and Flache, A. 2002. Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences* 99(suppl 3):7229–7236.

Macy, M. W. 1991. Learning to cooperate: Stochastic and tacit collusion in social exchange. *American Journal of Sociology* 97(3):808–843.

Masuda, N., and Nakamura, M. 2011. Numerical analysis of a reinforcement learning model with the dynamic aspiration level in the iterated prisoner's dilemma. *Journal of theoretical biology* 278(1):55–62.

Nowak, M., and Sigmund, K. 1993. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* 364(6432):56–58.

Nowak, M. A.; Page, K. M.; and Sigmund, K. 2000. Fairness versus reason in the ultimatum game. *Science* 289(5485):1773–1775.

Oosterbeek, H.; Sloof, R.; and Van De Kuilen, G. 2004. Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics* 7(2):171–188.

Pezzulo, G.; Butz, M.; Castelfranchi, C.; and Falcone, R., eds. 2008. *The Challenge of Anticipation*. Springer.

Rand, D. G.; Tarnita, C. E.; Ohtsuki, H.; and Nowak, M. A. 2013. Evolution of fairness in the one-shot anonymous ultimatum game. *Proc. Natl. Acad. Sci. USA* 110:2581–2586.

Rosen, R. 1985. *Anticipatory systems : philosophical, mathematical, and methodological foundations*. Pergamon Press, Oxford.

Roth, A. E., and Erev, I. 1995. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and economic behavior* 8(1):164–212.

Slimani, I.; El Farissi, I.; and Achchab, S. Application of game theory and neural network to study the behavioral probabilities in supply chain. *Journal of Theoretical & Applied Information Technology* 82(3):411–416.

Taiji, M., and Ikegami, T. 1999. Dynamics of internal models in game players. *Physica D: Nonlinear Phenomena* 134(2):253–266.

Zisis, I.; Di Guida, S.; Han, T.; Kirchsteiger, G.; and Lenaerts, T. 2015. Generosity motivated by acceptance-evolutionary analysis of an anticipation game. *Scientific reports* 5.