2019 IEEE 58th Conference on Decision and Control (CDC)
Palais des Congrès et des Expositions Nice Acropolis
Nice, France, December 11-13, 2019

Avoiding Chatter in an Online Co-Learning Algorithm Predicting Human Intention

Carol Young Ningshi Yao Fumin Zhang

Abstract— **Chatter can happen when an online learning algorithm is used by a robot to predict human intention while interacting with a human subject. When chatter happens, the learning algorithm continually changes its prediction, without reaching a constant prediction of human intention. Using the Rescorla-Wagner model for human learning, we analyze an expert based online learning algorithm and identify an invariant set in the state and parameter space where chatter will occur. Based on the chatter analysis, we also propose an improved expert based learning algorithm where the invariant set does not exist so that chatter can be avoided.**

I. Introduction

Co-learning occurs when two learning systems are considered as a single non-linear system with its own system dynamics. These co-learning systems are important to consider as the amount of learning robots increases. Some human-robot co-learning systems have been studied through experiments, e.g., a robot learns the type of humans that it is interacting with by using trained examples of humans interaction to select the optimal strategy [1], [2]. Or by jointly training reinforcement strategies in a scholastic game framework, a robot can use an optimally learned strategy to interact with a human [3].

A co-learning system can demonstrate *chatter*, which is a specific limit cycle phenomena where two learners both adjust their behaviors to correct for an error caused by a mismatch of their behaviors. This correction is simultaneous, but both new behaviors are still incompatible, which produce an error that leads to another simultaneous correction. This process repeats indefinitely without settling on a steady state. This dynamic behavior has been observed in human-human co-learning systems. In pedestrian counterflow studies, when two pedestrians meet, they may block each other repeatedly while trying to pass [4]. There exists evidence showing that human behavior will shift in response to robot behaviors, similar to the behavior shift in response to another human's behaviors [5], [6]. Therefore, chatter may exist in human-robot co-learning systems and should be analyzed.

This paper utilizes online learning algorithms for robots in a human-robot co-learning setting. An online learning algorithm learns a concept, which is how a human expects a robot to behave in a given situation. The algorithm takes in information about the situation, e.g., encountering a human, and produces an output, e.g., a behavior that a robot can perform. The algorithm then receives a feedback on its behavior. This set of information, output, and feedback is processed sequentially, and each set is only processed once. Thus the online leaning algorithm can be implemented without any pre-training, requiring less memory and processing time, while adjusting its behavior after every interaction [7].

Online learning can serve as a powerful tool in human robot interaction, especially when the two limitations, long learning time and unpredictable outputs, can be accounted for [8]. This can be achieved with communication cues that are provided from the robot to the human [9]. This can also be achieved by designing online learning algorithms to balance adaptiveness and consistency as shown by our previous research [10] which presented an online learning algorithm called Dual Expert Algorithm (DEA). DEA can be analyzed using Markov Chains, producing provable analytical performance bounds on adaptiveness and consistency.

This paper significantly extends our previous work where human intention was treated as random drift [10], by considering the fact that a human can also learn from a robot. In this paper, we model the human learning behavior using a well accepted and widely used model, called Resorla-Wagner model [11], to describe how a human updates the internal expectation. The DEA is used as the learning algorithm for robots. We show that chatter can occur in this system and we improve DEA to eliminate the chatter. The contributions of this paper are as follows. (1) We discover the dynamical model which triggers chatter during co-learning. We model the co-learning system as a double feedback system with binary outputs. The state space of this co-learning system consists of an internal state of the human and the states of the DEA Markov Chain. This co-learning model allows us to compute an invariant set in the state space that leads to chatter. (2) We improve the DEA and present a new algorithm, called Human Aware Dual Expert Algorithm (HADEA), which can avoid chatter during co-learning. We rigorously prove that using HADEA, there is no chatter zone (e.g. an invariant set in the state space of the co-learning system). The results are justified using numerical methods in simulations.

II. Problem Setup

In our previous work, we assumed that the expectation of a human subject is only affected by drift and independent from the performance of the learning algorithm used by the robot. This assumption usually does not hold because the

*The research work is supported by ONR grants N00014-19-1-2266 and N00014-16-1-2667; NSF grants OCE-1559475, CNS-1828678, and S&AS-1849228; NRL grants N0017317-1-G001 and N00173-19-P-1412; NOAA NA16NOS0120028; an ARCS and a CSC Scholar Award.

Carol Young, Ningshi Yao and Fumin Zhang are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332, USA {cyoung44,nyao6,fumin}@gatech.edu.

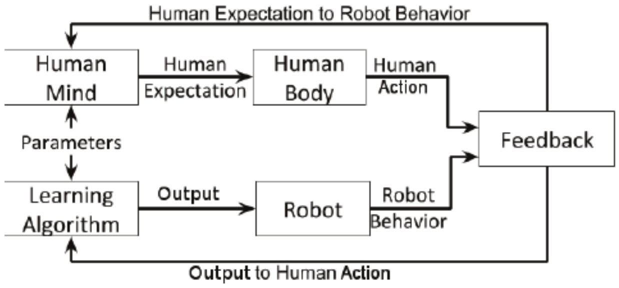

Fig. 1. Layout of a co-learning system consisting of a human and a robot with a learning algorithm

human's expectation of the robot can change due to the changes in robot's behavior. Therefore, we study the mutual influence between the changing expectation of a human and the learning algorithm used by a robot. In this paper, we only consider a simple case where both the human and the robot have only two possible actions, but it will illustrate the key idea of how to analyze the co-learning system behavior when chatter occurs and how to solve the chatter issue. The actions for human is denoted by A_t which represents the action taken by the human at t-th interaction. It can have only two values, -1 or 1. For the robot, the actions it can take are denoted by λ_t, representing the action taken at t-th interaction, which also has the value of -1 or 1. If $A_t = \lambda_t$, then we say the action of the human matches the action of the robot. Otherwise, the two actions do not match.

We model a co-learning system using two feedback loops, as shown by Figure 1. For each human robot interaction, the human has an expectation of what action the robot will take for the upcoming interaction. Based on the expectation, the human takes one corresponding action out of two possible human actions. A feedback is given to the human by comparing the robot's action to the human's action. If they match with each other, then the feedback is positive. Otherwise, the feedback is negative. Based on the feedback i.e. whether a feedback is positive or not, we assume the human will adjust its internal states (characterized by some internal parameters) and may change the expectation for the next interaction. Similar to the human's learning process, the learning algorithm implemented on a robot predicts an output based on its learned concept of human. The output selects one robot action out of two. The feedback is also given to the learning algorithm after an interaction to adjust its internal parameters in order to make a correct prediction for the next interaction. The whole co-learning system can be viewed as a closed loop system and we can analyze its dynamic behavior. In this section, we first introduce the models of blocks "human mind" and "learning algorithm" in Figure 1.

A. Human Learning Model

To model the human mind, we use a well studied psychological model called the Rescorla-Wagner model [11], which has been used to describe the human learning process that associates a conditioned stimuli with an unconditioned stimuli. We consider the arrival of a robot as the conditioned stimuli and the behavior of the robot as the unconditioned stimuli. Then the Rescorla-Wagner model describes how a human learns to predict the robot's behavior. The Rescorla-Wagner model is described by

$$V_t = (1-\gamma)V_{t-1} + \gamma\lambda_{t-1} \quad (1)$$

where γ is a constant with real value in $(0,1)$ representing the weight the human puts on new information and V_t is a real number in the range of $[-1,1]$, representing the internal state of the human at the t-th interaction. We denote V_0 as the initial value of V_t before the human starts interacting with the robot, which is defined by the Rescorla-Wagner model as being in the range of $[-1,1]$. Based on the initial value of V_0, we can show the range of V_t in the following lemma.

Lemma 2.1: $V_t \in [-1,1]$ for all $t > 0$.

Proof: We will prove this lemma by induction. Assume that $V_{t-1} \in [-1,1]$. Since $-1 \le V_{t-1} \le 1$, then we have

$$-1(1-\gamma)+\gamma\lambda_{t-1} \le (1-\gamma)V_{t-1}+\gamma\lambda_{t-1} \le 1(1-\gamma)+\gamma\lambda_{t-1}. \quad (2)$$

Based on equation (1), we can replace $(1-\gamma)V_{t-1}+\gamma\lambda_{t-1}$ with V_t and rearrange (2), which results in

$$-1+\gamma(\lambda_{t-1}+1) \le V_t \le 1+\gamma(\lambda_{t-1}-1).$$

By definition, the value of λ_{t-1} can only be either -1 or 1. The possible values for the lower bound $-1+\gamma(\lambda_{t-1}+1)$ are -1 or $-1+2\gamma$. And the possible values for the upper bound $1+\gamma(\lambda_{t-1}-1)$ are $1-2\gamma$ and 1. Since $\gamma \in [0,1]$, we have $-1+2\gamma \in [-1,1]$, which means that the smallest possible value for the lower bound of V_t is -1. Likewise, the upper bound $1-2\gamma$ is also in the range of $[-1,1]$. The largest possible value for the upper bound of V_t is 1. Therefore, $V_t \in [-1,1]$ if $V_{t-1} \in [-1,1]$. Since the initial value $V_0 \in [-1,1]$, then by induction, $V_t \in [-1,1]$, for any $t > 1$. □

Once we have the internal state V_t of the human, the human action A_t can be determined based on the value of V_t

$$A_t = \begin{cases} 1 & \text{if } V_t \ge 0 \\ -1 & \text{if } V_t < 0 \end{cases} \quad (3)$$

The human subject is only allowed to make two possible choices of the actions represented by the two values of A_t. Correspondingly, our learning algorithm will control the robot to produce two reactions respectively.

B. Dual Expert Algorithm

For the "learning algorithm" block in Figure 1, we use DEA in [10] as the learning algorithm for a robot. Algorithm 1 presents the pseudo code of DEA. In DEA, each expert represents an action and is assigned a weight, W_{-1} and W_1. Line 2 indicates that DEA selects the index of the expert with the highest weight. If the two experts have the same weights at an iteration, then DEA selects -1. The two weights are adjusted after DEA made a selection of action, and gets a feedback after comparing the action with human action. If the feedback is negative, i.e. the actions of the human and robot do not match, then the weight associated with the selected action will be half of its previous value. If the feedback is positive, i.e. the actions of the human and robot match, and the weight of the selected expert is less than 0.5, then

Algorithm 1 Dual Expert Algorithm

1: Set $W_1 = W_{-1} = 0.5$
2: Choose selection $\lambda_t = \min(\text{argmax}\{W_1, W_{-1}\})$
3: **if** Error ($\lambda_t = -A_t$) **then**
4: $\quad W_{\lambda_t} = \frac{W_{\lambda_t}}{2}$
5: **else if** Correct ($\lambda_t = A_t$) **then**
6: $\quad$ **if** $W_{\lambda_t} < 0.5$ **then**
7: $\quad\quad W_{\lambda_t} = 2W_{\lambda_t}$
8: $\quad$ **else**
9: $\quad\quad W_{\lambda_t} = W_{\lambda_t}$

the weight is multiplied by 2. Otherwise, the weight of the selected expert remains the same. The algorithm will select the next action by comparing the two weights, and the action associated with the larger weight will always be selected. The actions taken is determined by the weights of DEA

$$\lambda_t = \min(\text{argmax}\{W_1, W_{-1}\}) \tag{4}$$

where W_{-1} and W_1 are the weights of the experts in DEA at the t-th interation. Note λ_t is also the index of the winning expert. We introduce two additional variables which will be used in this analysis. The first variable R_t is defined as

$$R_t = \begin{cases} \log_2\left(\frac{2W_1}{W_{-1}}\right) & \text{if } W_1 \geq W_{-1} \\ \log_2\left(\frac{W_{-1}}{W_1}\right) & \text{if } W_1 < W_{-1} \end{cases} \tag{5}$$

which represents the minimum number of errors that must occur before the DEA switches the selection of experts, meaning changing from one action to the another action. The second variable s_t is defined as

$$s_t = \begin{cases} \log_2\left(\frac{1}{W_{-1}}\right) & \text{if } W_1 \geq W_{-1} \\ \log_2\left(\frac{0.5}{W_1}\right) & \text{if } W_1 < W_{-1} \end{cases} \tag{6}$$

Here s_t represents the maximal achievable value of R_t. Because the weights in DEA have an upper bound 0.5, replacing W_1 and W_{-1} in (5) with 0.5 results in (6). Based on this definition, $R_t \leq s_t$ always holds.

The three variables, (λ_t, R_t, s_t) is used to define a Markov chain of DEA as shown in Figure 2. And based on line 1 of Algorithm 1, the initially selected action is -1. Therefore, the initial state of DEA is $(-1,1,1)$. The updating rules for each variable are as follows

$$\lambda_t = \begin{cases} -\lambda_{t-1} & \text{if } \lambda_{t-1} \neq A_t \text{ and } R_{t-1} = 1 \\ \lambda_{t-1} & \text{otherwise} \end{cases} \tag{7}$$

$$R_t = \begin{cases} R_{t-1}+1 & \text{if } \lambda_{t-1} = A_{t-1} \text{ and } R_{t-1} < s_{t-1} \\ R_{t-1} & \text{if } \lambda_{t-1} = A_{t-1} \text{ and } R_{t-1} = s_{t-1} \\ R_{t-1}-1 & \text{if } \lambda_{t-1} \neq A_{t-1} \text{ and } R_{t-1} > 1 \\ R_{t-1} & \text{if } \lambda_{t-1} \neq A_{t-1} \text{ and } R_{t-1} = 1 \end{cases} \tag{8}$$

$$s_t = \begin{cases} s_{t-1}+1 & \text{if } \lambda_{t-1} = -1 \neq A_t \text{ and } R_{t-1} = 1 \\ s_{t-1} & \text{otherwise} \end{cases} \tag{9}$$

We assume that the robot enacts a control law enabling it to perform behavior λ_t, and has detection capabilities to detect the human action A_t. It is obvious that the robot cannot directly observe the human's internal parameters γ or internal state V_t, but it can observe the human's action A_t.

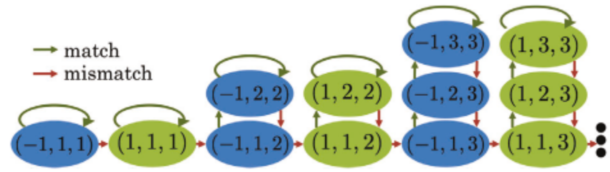

Fig. 2. DEA automata with each state defined in terms of (λ_t, R_t, s_t). The green arrows represent the cases where the actions of human and robot match. The red arrows represent the cases where the actions of human and robot do not match each other.

The behavior of the robot can influence the human's internal state V_t but it cannot change the parameter γ. That is to say, we assume that γ remains as a constant during co-learning.

C. Chatter definition

Based on the mathematical models of human learning and DEA. The chatter behavior in the co-learning system with binary actions can be defined as follows.

Definition 2.1: If there exist a t_0 such that $A_t = -A_{t-1}$, $\lambda_t = -\lambda_{t-1}$, and $A_t = -\lambda_t$ hold for all $t > t_0$, then we say a *chatter* occurs in the co-learning system with binary actions.

Since we only allow binary values for A_t and λ_t, the chatter behavior in co-learning is a limit cycle in the dynamics. This chatter is not desired for co-learning. Therefore, we will analyze the co-learning system to determine the condition when chatter occurs. And based on this analysis, we modify DEA in order to prevent chatter.

III. Chatter Analysis

The human is described by two internal parameters V_t and γ, where V_t is the state that changes over time. We define the set $B_{A_t,n}$ to be the set of all possible values of (γ, V_t) that produce output A_t and takes a minimum of n errors from time t to produce the output $-A_t$. This set is important for chatter analysis since any human internal state V_t starting in this set will generate chatter eventually. We will compute this set in different settings.

Lemma 3.1: For $A_t = -1$ and $n = 1$,

$$B_{-1,1} = \left\{(V_t,\gamma) | \gamma \in (0,1), V_t \in \left[\frac{-1}{1-\gamma}+1, 0\right) \bigcap [-1,0)\right\}.$$

Proof: In Lemma 2.1, we have already shown that $V_t \in [-1,1]$ holds for all $t > 0$. Here, we will show that in the set $B_{-1,1}$, V_t also needs to satisfy $V_t \in \left[\frac{-1}{1-\gamma}+1, 0\right)$.

Based on definition, the set $B_{-1,1}$ produces output $A_t = -1$, which means the corresponding internal state of human V_t must satisfy $V_t < 0$ according to (3). And $n = 1$ means it requires 1 mismatch for the human to switch to action 1. That is to say, the action of the robot λ_t at iteration t is not the same as A_t, i.e., $\lambda_t = 1$ and at the next iteration, the action of the human changes to be 1, i.e., $A_{t+1} = 1$. Therefore, we have $V_{t+1} \geq 0$. By (1), we have $V_{t+1} = (1-\gamma)V_t + \gamma\lambda_t$. With $\lambda_t = 1$, we have

$$V_{t+1} = (1-\gamma)V_t + \gamma \geq 0 \Rightarrow (1-\gamma)V_t \geq -\gamma. \tag{10}$$

Since $\gamma \in (0,1)$, $1-\gamma > 0$. Then we can divide $1-\gamma$ on the both sides of (10), which results in

$$(1-\gamma)V_t \geq -\gamma \Rightarrow V_t \geq \frac{-\gamma}{1-\gamma} = \frac{-1}{1-\gamma}+1. \quad \square$$

Lemma 3.2: For an arbitrary number of mismatches $n \geq 2$,

$$B_{-1,n} = \Big\{(V_t,\gamma) | \gamma \in (0,1), \quad V_t \in \Big[\frac{-1}{(1-\gamma)^n}+1, \frac{-1}{(1-\gamma)^{n-1}}+1\Big) \bigcap [-1,0)\Big\}.$$

Proof: Same as the the proof of Lemma 3.1, since $A_t = -1$, we have $V_t < 0$. And since it requires minimal n mismatches for the human to switch to action 1, we have $V_t = V_{t+1} = \cdots = V_{t+n-1} < 0$ and $V_{t+n} \geq 0$. Additionally, $\lambda_0 = \cdots = \lambda_{n-1} = 1$. Utilizing the result from Lemma 3.1, we can see that V_{t+n-1} must satisfy $\frac{-1}{1-\gamma}+1 \leq V_{t+n-1} < 0$. With $V_{t+n-1} = (1-\gamma)V_{t+n-2} + \gamma\lambda_{t+n-2}$ and $\lambda_{t+n-2} = 1$, we can derive

$$\frac{-1}{(1-\gamma)^2} + \frac{1-\gamma}{1-\gamma} \leq V_{t+n-2} < \frac{-1}{(1-\gamma)} + 1. \qquad (11)$$

Replacing V_{t+n-2} with $(1-\gamma)V_{t+n-3} + \gamma$ in (11), we have $\frac{-1}{(1-\gamma)^3} + 1 \leq V_{t+n-3} < \frac{-1}{(1-\gamma)^2} + 1$. Then repeat the above process, we have $\frac{-1}{(1-\gamma)^n} + 1 \leq V_t < \frac{-1}{(1-\gamma)^{n-1}} + 1$. □.

By symmetric arguments, we can also conclude that the following Lemmas hold.

Lemma 3.3: For $A_t = 1$ and $n = 1$,

$$B_{1,1} = \Big\{(V_t,\gamma) | \gamma \in (0,1), V_t \in \Big[0, \frac{1}{1-\gamma} - 1\Big) \bigcap [0,1]\Big\}.$$

Lemma 3.4: For an arbitrary number of mismatches $n \geq 2$,

$$B_{1,n} = \Big\{(V_t,\gamma) | \gamma \in (0,1), \quad V_t \in \Big[\frac{1}{(1-\gamma)^{n-1}}-1, \frac{1}{(1-\gamma)^{n}}-1\Big) \bigcap [0,1]\Big\}.$$

After obtaining the representation of the sets, we now show that there always exist γ and V_t that are in these sets, i.e., these sets are not empty.

Lemma 3.5: For any finite n, $B_{A_t,n}$ is not empty.

Proof: If $A_t = 1$ or -1 and $n = 1$, since $0 < \gamma < 1$, we have $0 < 1-\gamma < 1$ which results in $\frac{-1}{1-\gamma} < -1$ and $\frac{1}{1-\gamma} > 1$. Then we have $\frac{-1}{1-\gamma} + 1 < 0$ and $\frac{1}{1-\gamma} - 1 > 0$ is always satisfied. Therefore, sets $B_{1,1}$ and $B_{-1,1}$ in Lemmas 3.1 and 3.3 are not empty.

For $A_t = -1$ and $n \geq 2$, we can first show the interval $\Big[\frac{-1}{(1-\gamma)^n}+1, \frac{-1}{(1-\gamma)^{n-1}}+1\Big)$ is not empty by $\frac{-1}{(1-\gamma)^{n-1}}+1 - \Big(\frac{-1}{(1-\gamma)^n}+1\Big) = \frac{\gamma}{(1-\gamma)^n} > 0$ because $\gamma > 0$ and $1-\gamma > 0$. Then we will show that there always exists a γ so that $\frac{-1}{(1-\gamma)^{n-1}}+1 > -1$, i.e., the intersection of the two interval for V_t is not an empty set. Let $0 < \gamma < 1 - (\frac{1}{2})^{\frac{1}{n-1}}$. Note that $0 < (\frac{1}{2})^{\frac{1}{n-1}} < 1$ for all $n \geq 2$ so $0 < 1 - (\frac{1}{2})^{\frac{1}{n-1}} < 1$. We have

$$1-\gamma > \Big(\frac{1}{2}\Big)^{\frac{1}{n-1}} \Rightarrow (1-\gamma)^{n-1} > \frac{1}{2} \Rightarrow \frac{1}{(1-\gamma)^{n-1}} < 2. \qquad (12)$$

Then we can derive that

$$\frac{-1}{(1-\gamma)^{n-1}} > -2 \Rightarrow \frac{-1}{(1-\gamma)^{n-1}} + 1 > -1.$$

Therefore, the set $B_{-1,n}$ is not empty for $n \geq 2$.

Likewise, for $A_t = 1$ and $n \geq 2$, we have

$$\frac{1}{(1-\gamma)^n} - 1 - \Big(\frac{1}{(1-\gamma)^{n-1}} - 1\Big) = \frac{\gamma}{(1-\gamma)^n} > 0.$$

And if $0 < \gamma < 1 - (\frac{1}{2})^{\frac{1}{n-1}}$, based on (12), we have

$$\frac{1}{(1-\gamma)^{n-1}} < 2 \Rightarrow \frac{1}{(1-\gamma)^{n-1}} - 1 < 1.$$

Therefore, the set $B_{1,n}$ is not empty for $n \geq 2$. □

Next, we show that a mismatch at $(t-1)$-th interaction will cause a changed output at t-th iteration if the human internal parameters (γ, V_{t-1}) are in $B_{A_{t-1},1}$ at $(t-1)$-th interaction.

Lemma 3.6: If $(\gamma, V_{t-1}) \in B_{A_{t-1},1}$ and one mismatch at $(t-1)$-th interaction, then $(\gamma, V_t) \in B_{-A_{t-1},1}$ at t-th iteration.

Proof: We will prove this Lemma by two cases.

Case 1: $A_{t-1} = -1$. Since $(\gamma, V_{t-1}) \in B_{-1,1}$, we have $\frac{-1}{1-\gamma} + 1 \leq V_{t-1} < 0$ based on Lemma 3.1. And a mismatch occurs at $(t-1)$-th iteration meaning that $\lambda_{t-1} \neq A_{t-1}$, i.e., $\lambda_{t-1} = 1$. And based (1) and $\lambda_{t-1} = 1$, we have $V_t = (1-\gamma)V_{t-1} + \gamma$. Since $1-\gamma > 0$, we have

$$V_t = (1-\gamma)V_{t-1} + \gamma \geq (1-\gamma) \cdot \Big(\frac{-1}{1-\gamma} + 1\Big) + \gamma = 0 \text{ and}$$
$$V_t = (1-\gamma)V_{t-1} - \gamma < (1-\gamma) \cdot 0 + \gamma = \gamma.$$

Therefore, $V_t \in [0, \gamma)$. And we compare γ with the upper bound of V_t in set $B_{1,1}$ by taking the difference $\gamma - \Big(\frac{1}{1-\gamma} - 1\Big) = \frac{-\gamma^2}{1-\gamma} < 0$ which means that $[0,\gamma) \subset B_{1,1}$. Therefore, V_t is in set $B_{1,1}$.

Case 2: $A_{t-1} = 1$. Since $(\gamma, V_{t-1}) \in B_{1,1}$, we have $0 \leq V_{t-1} < \frac{1}{1-\gamma} - 1$ based on Lemma 3.3. And a mismatch occurs at $(t-1)$-th iteration meaning that $\lambda_{t-1} = -1$. And based (1) and $\lambda_{t-1} = -1$, we have $V_t = (1-\gamma)V_{t-1} - \gamma$. Since $1-\gamma > 0$, we have

$$V_t = (1-\gamma)V_{t-1} - \gamma \geq (1-\gamma) \cdot 0 - \gamma = -\gamma \text{ and}$$
$$V_t = (1-\gamma)V_{t-1} - \gamma < (1-\gamma) \cdot \Big(\frac{1}{1-\gamma} - 1\Big) - \gamma = 0.$$

Therefore, $V_t \in [-\gamma, 0)$. And we compare $-\gamma$ with the lower bound of V_t in set $B_{-1,1}$ by taking the difference $-\gamma - \Big(\frac{-1}{1-\gamma} + 1\Big) = \frac{\gamma^2}{1-\gamma} > 0$ which means that V_t is within the set $B_{-1,1}$. □

We can then give a sufficient condition for chatter to occur if the robot used DEA as its learning algorithm.

Lemma 3.7: For the DEA, if $\lambda_{t-1} = -A_{t-1}$, $R_{t-1} = n_{t-1} = 1$ and $(\gamma, V_{t-1}) \in B_{A_{t-1},1}$, then a chatter will occur.

Proof: Since $\lambda_{t-1} = -A_{t-1}$ and $R_{t-1} = 1$, based on the 4-th condition in (8), $R_t = 1$. And by (7), $\lambda_t = -\lambda_{t-1}$.

According to Lemma 3.6, if the state at the $(t-1)$-th iteration is within the set $B_{A_{t-1},1}$ and $\lambda_{t-1} = -A_{t-1}$, then $(\gamma, V_t) \in B_{-A_{t-1},1}$ meaning $A_t = -A_{t-1}$. This means that at the t-th iteration, $A_t = -A_{t-1} = \lambda_{t-1} = -\lambda_t$, i.e. a mismatch occurs at the t-th iteration. Then the conditions in Lemma 3.6 are also satisfied the t-th iteration, which results in a mismatch at the $(t+1)$-th iteration. Thus for all future iterations, mismatches will keep occurring. This leads to chatter according to Definition 2.1. □

We can conclude that chatter occurs if the initial values for the human's internal state V_0 satisfy the following condition.

Theorem 3.8: For DEA, if (γ, V_0) is in set $B_{-1,1}$, then chatter will occur.

Algorithm 2 Human Aware Dual Expert Algorithm Addition

1: **if** $\lambda_t \neq \Lambda = \min(\text{argmax}\{W_{-1}, W_1\})$ **then**

2: $\quad W_\Lambda = \begin{cases} 2W_\Lambda & \text{if } W_\Lambda < 0.5 \\ 0.5 & \text{otherwise} \end{cases}$

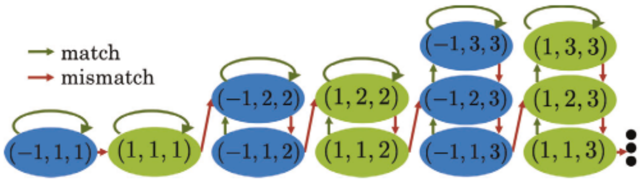

Fig. 3. The automata of HADEA with each state defined as (λ_t, R_t, s_t)

Proof: The initial state of DEA is given as $R_0 = s_0 = \lambda_0 = 1$. And since $V_0 < 0$, we have $A_0 = -1$ which means $\lambda_0 = -A_0$. Then by Lemma 3.7, the system will produce chatter. □

IV. Human Aware DEA

In this section, we present a revised DEA which can prevent chatter during the co-learning process. The algorithm is called Human Aware DEA (HADEA). Algorithm 2 presents the pseudo code which we add between line 4 and line 5 in algorithm 1 to create HADEA. The meaning of these lines are that, if the learning algorithm will change the index of selected expert in the next iteration, i.e. $\lambda_t \neq \Lambda$, then an additional weight increase may be performed at the current iteration. For example, the initial values of HADEA are $W_{-1} = W_1 = 0.5$ and the action is chosen to be $\lambda_1 = -1$. If the action of the robot matches with the human's action, the weights remain the same and the algorithm goes to the second iteration. But if actions do not match, i.e. $\lambda_1 = -A_1$, then the weight W_{-1} of expert representing action -1 is reduced to 0.25. Using the updated weight, we have $W_1 = 0.5 > W_{-1}$. Therefore, $\Lambda = 1$ and $\lambda_1 \neq \Lambda$. And the algorithm goes to line 6. Since $W_1 = 0.5$, it remains to be 0.5 after running line 6 and the algorithm goes to the second iteration. At the second iteration, by line 2, $\lambda_2 = 1$. If an error occurs, then W_1 is decreased to be 0.25. Then using the updated weights, we have $W_1 = 0.25 = W_{-1}$. Therefore, $\Lambda = -1$ and $\lambda_2 \neq \Lambda$. In this case, W_{-1} is increased to 0.5. This is when an additional weight increase is performed.

This new updating rule generates a different automata of HADEA compared to DEA, which is shown by Figure 3. This new algorithm produces the same updating equations for λ_t by (7), and s_t by (9). However, the updating equation for R_t has changed as

$$R_t = \begin{cases} R_{t-1}+1 & \text{if } \lambda_{t-1} = A_{t-1} \text{ and } R_{t-1} < s_{t-1} \\ R_{t-1} & \text{if } \lambda_{t-1} = A_{t-1} \text{ and } R_{t-1} = s_{t-1} \\ R_{t-1}-1 & \text{if } \lambda_{t-1} \neq A_{t-1} \text{ and } R_{t-1} > 0 \\ 1 & \text{if } \lambda_{t-1} \neq A_{t-1}, R_{t-1} = 0 \text{ and } s_t = 1 \\ 2 & \text{if } \lambda_{t-1} \neq A_{t-1}, R_{t-1} = 0 \text{ and } s_t > 1 \end{cases} \quad (13)$$

In order to show that HADEA does not generate chatter, we show that all possible values of human's internal states (γ, V_t) do not lead to chatter.

Theorem 4.1: If $A_t = \lambda_t$, then there is no chatter after t.

Proof: If $A_t = \lambda_t$, then from (1) and (3), $A_{t+1} = A_t$. Because $A_t = \lambda_t = -1$ means $V_t < 0$, then $V_{t+1} = (1-\gamma)V_t - \gamma$. With $1-\gamma > 0$, $V_t < 0$ and $-\gamma < 0$, we have $V_{t+1} < 0$ and $A_{t+1} = -1 = A_t$. Similarly, if $A_t = \lambda_t = 1$, then $V_{t+1} = (1-\gamma)V_t + \gamma > 0$ which results in $A_{t+1} = 1 = A_t$. From (7), $\lambda_{t+1} = \lambda_t$. Therefore, $A_{t+1} = \lambda_{t+1}$. Repeat the above argument, we show that there will be no chatter after t if $A_t = \lambda_t$. □

Lemma 4.2: For a set B_{A_t,n_t}, if $n_t \geq 2$ and $\lambda_t \neq A_t$, then the set $B_{A_{t+1},n_{t+1}}$ is identical to the set B_{A_t,n_t-1}

Proof: If $n_t \geq 2$ and $A_t = -1$, from Lemma 3.2, we have $V_t \in \left[\frac{-1}{(1-\gamma)^{n_t}} + 1, \frac{-1}{(1-\gamma)^{n_t-1}} + 1\right)$. Using (1) and $\lambda_t = 1$, yields

$$V_{t+1} = (1-\gamma)V_t + \gamma \geq (1-\gamma) \cdot \left[\frac{-1}{(1-\gamma)^{n_t}} + 1\right] + \gamma \text{ and}$$

$$V_{t+1} = (1-\gamma)V_t + \gamma < (1-\gamma) \cdot \left[\frac{-1}{(1-\gamma)^{n_t-1}} + 1\right] + \gamma. \quad (14)$$

Reorganizing (14) gives $V_{t+1} \in \left[\frac{-1}{(1-\gamma)^{n_t-1}} + 1, \frac{-1}{(1-\gamma)^{n_t-2}} + 1\right)$ which would be in set B_{-1,n_t-1}. Likewise, for $A_t = 1$, from Lemma 3.4, we have $V_t \in \left[\frac{1}{(1-\gamma)^{n_t-1}} - 1, \frac{1}{(1-\gamma)^{n_t}} - 1\right)$. Using (1) and $\lambda_t = -1$, we find the range for V_{t+1} is in $\left[\frac{1}{(1-\gamma)^{n_t-2}} - 1, \frac{1}{(1-\gamma)^{n_t-1}} - 1\right)$ which would be in set B_{1,n_t-1}.

Therefore, after an error occurs at t-th iteration, if $n_t \geq 2$, then $B_{A_{t+1},n_{t+1}} = B_{A_t,n_t-1}$. □

Theorem 4.3: If $A_t = -\lambda_t$ and $R_t \neq n_t$, then there is no chatter after t.

Proof: First we consider $R_t > n_t$. Since $A_t \neq \lambda_t$, an error occurs at time t. This can lead to two cases.

If $n_t = 1$, then by Lemma 3.6, we have $(\gamma, V_t) \in B_{A_t,1}$ and $(\gamma, V_{t+1}) \in B_{-A_t,1}$ meaning that $A_{t+1} = -A_t$. And since $R_t > n_t = 1$, the HADEA updating equations (7) and (13) will give the result that $R_{t+1} = R_t - 1$ and $\lambda_{t+1} = \lambda_t$. Since $\lambda_{t+1} = A_{t+1}$, by Theorem 4.1, there is no chatter in the future.

If $n_t \geq 2$, then by Lemma 4.2, $A_{t+1} = A_t$, and $n_{t+1} = n_t - 1$. And since $R_t > n_t > 1$, then based on the HADEA updating rules (7) and (13), we have $R_{t+1} = R_t - 1$ and $\lambda_{t+1} = \lambda_t$. Then we have $R_{t+1} = R_t - 1 > n_{t+1} = n_t - 1$ and $A_{t+1} \neq \lambda_{t+1}$ where the only changing variable is n_t. After $n_t - 1$ iterations, we will have $R_{t+n_t-1} > n_{t+n_t-1} = 1$, which is the case where $n_t = 1$. Therefore, after $(t + n_t - 1)$ iterations, there is no chatter.

Now we consider $n_t > R_t$. Since $A_t \neq \lambda_t$, an error occurs at iteration t. This can lead to two distinct cases depending on if $R_t = 1$ or $R_t > 1$.

If $R_t = 1$, then by (7), we have $\lambda_{t+1} = -\lambda_t$. And since $n_t > R_t = 1$ by Lemma 4.2, we know $A_{t+1} = A_t$. Then we will have $\lambda_{t+1} = A_{t+1}$. By Theorem 4.1 there is no chatter.

If $R_t > 1$, then by the HADEA updating equations (7) and (13), we have $R_{t+1} = R_t - 1$ and $\lambda_{t+1} = \lambda_t$. And by Lemma 4.2, we can derive that $A_{t+1} = A_t$ and $n_{t+1} = n_t - 1$. Therefore, we have $n_{t+1} = n_t - 1 > R_{t+1} = R_t - 1$ and $A_{t+1} \neq \lambda_{t+1}$ after an iteration with both n_t and R_t decreased by 1. By induction, after a total of $\Delta t = R_t - 1$ iterations, $n_{t+\Delta t} > R_{t+\Delta t} = 1$, which as shown above for the case of $R_t = 1$.

In summary, all possible situations that satisfy $A_t = -\lambda_t$ and $R_t \neq n_t$ have been discussed. None of these situations lead to chatter. Therefore, we can conclude that if $A_t = -\lambda_t$ and $R_t \neq n_t$, then there is no chatter. □

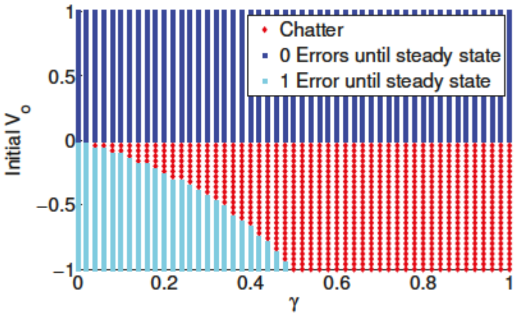

Fig. 4. Simulation results for DEA based off of γ and V_0.

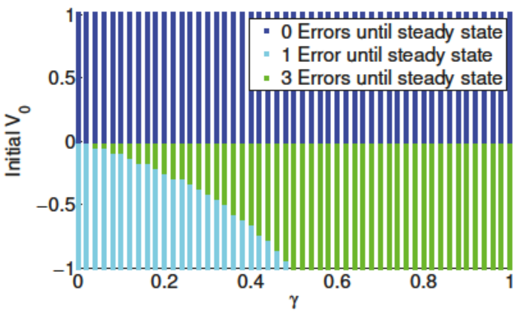

Fig. 5. Simulation results for HADEA based off of γ and V_0.

Theorem 4.4: If $A_t \neq \lambda_t$ and $R_t = n_t$, then there is no chatter when using HADEA.

Proof: If $A_t \neq \lambda_t$ and $R_t = n_t$, then using equations (7), (13), and Lemma 4.2, we have $R_{t+1} = R_t - 1$ and $n_{t+1} = n_t - 1$. After $\Delta t = R_t - 1$ iterations, $R_{t+\Delta t} = n_{t+\Delta t} = 1$ and $A_{t+\Delta t} = -\lambda_{t+\Delta t}$. By Lemma 3.6, $A_{t+\Delta t+1} = -A_{t+\Delta t}$ and $n_{t+\Delta t+1} = 1$.

The following cases must be considered to show that no chatter happens. The first case is if $s_{t+\Delta t+1} > 1$. The second case is if $s_{t+\Delta t+1} = 1$.

If $s_{t+\Delta t+1} > 1$then by (13) and (7), we have $R_{t+\Delta t+1} = 2$ and $\lambda_{t+\Delta t+1} = -\lambda_{t+\Delta t}$. Then we will have the condition $R_{t+\Delta t+1} \neq n_{t+\Delta t+1}$ and $\lambda_{t+\Delta t+1} = -A_{t+\Delta t+1}$. By Theorem 4.3, there will be no chatter in the future.

If $s_{t+\Delta t+1} = 1$, then by (13), $R_{t+\Delta t+1} = 1$ and $\lambda_{t+\Delta t+1} = \lambda_{t+\Delta t}$. In the following, we will discuss two cases $\lambda_{t+\Delta t+1} = -1$ and $\lambda_{t+\Delta t+1} = 1$.

If $\lambda_{t+\Delta t+1} = -1$, then by (7), (8), (9), we have $s_{t+\Delta t+2} = 2$, $R_{t+\Delta t+2} = 2$ and $\lambda_{t+\Delta t+2} = -\lambda_{t+\Delta t+1}$, And by Lemma 3.6, $A_{t+\Delta t+2} = -A_{t+\Delta t+1}$ and $n_{t+\Delta t+2} = 1$. Then we have $R_{t+\Delta t+2} \neq n_{t+\Delta t+2}$ and $\lambda_{t+\Delta t+2} = -A_{t+\Delta t+2}$. By Theorem 4.3, there will be no chatter in the future.

If $\lambda_{t+\Delta t+1} = 1$, then $s_{t+\Delta t+2} = 1$, $R_{t+\Delta t+2} = 1$ and $\lambda_{t+\Delta t+2} = -\lambda_{t+\Delta t+1}$. By Lemma 3.6, $A_{t+\Delta t+2} = -A_{t+\Delta t+1}$ and $n_{t+\Delta t+2} = 1$. This is now the same as the previous case, $\lambda_{t+\Delta t+1} = -1$, and so this case also does not produce chatter.

Since all the cases have been discussed, all of them do not produce chatter. Therefore, if $A_t \neq \lambda_t$ and $R_t = n_t$ then there is no chatter when using HADEA. □

Combining Theorem 4.1, Theorem 4.3, and Theorem 4.4 we conclude that there is no values of γ and V_t that can produce chatter for the HADEA.

V. Simulation

To support the chatter analysis for DEA and HADEA, simulations were run for both algorithms. Each algorithm was initialized with $R_0 = 1$, $s_0 = 1$ and $\lambda_0 = 1$, and was run using the Rescorla-Wagner model to model the human. Different values of (γ, V_0) of the Rescorla-Wagner model were sampled. For the sample of V_0, we divide the range $[-1, 1]$ into 50 even portions and take one value from each portion as the sampled value for V_0. For the sample of γ, we divide the range $(0, 1)$ into 50 even portions and take one value from each as the sampled value. The amount of iterations that an algorithm took to reach a steady state, i.e. $\lambda_t = A_t$, was recorded for each (γ, V_0) pair. If a steady state was not reached after 50 iterations, then the co-learning was considered to have a chatter give the sample pair (γ, V_0).

Figure 4 shows the simulation result for DEA. The red area indicates the values of γ and V_0 where chatter occurs. These data points match the set $B_{-1,1}$, which is consistent with our analysis. Figure 5 shows the steps required to reach steady state when HADEA is used. There is no chatter present e.g. no red region. The values in $B_{-1,1}$ only requiring 3 steps to reach the steady state, which matches with our analysis.

VI. Conclusion

In this paper, we have used the Rescorla-Wagner and DEA models to show that a human robot co-learning system can demonstrate chatter. Upon identifying chatter, we developed the HADEA and showed that HADEA can avoided chatter. Besides, HADEA has a drastically increased consistency over DEA, with only a limited decrease in adaptiveness.

References

[1] S. Nikolaidis, J. Forlizzi, D. Hsu, J. Shah, and S. Srinivasa, "Mathematical models of adaptation in human-robot collaboration," *arXiv preprint arXiv:1707.02586*, 2017.

[2] S. Nikolaidis, D. Hsu, and S. Srinivasa, "Human-robot mutual adaptation in collaborative tasks: Models and experiments," *The International Journal of Robotics Research*, vol. 36, no. 5-7, pp. 618–634, 2017.

[3] M. Barlier, J. Perolat, R. Laroche, and O. Pietquin, "Human-machine dialogue as a stochastic game," in *16th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL 2015)*, 2015.

[4] D. Helbing, P. Molnár, I. J. Farkas, and K. Bolay, "Self-organizing pedestrian movement," *Environment and planning B: planning and design*, vol. 28, no. 3, pp. 361–383, 2001.

[5] T. Nomura, T. Kanda, T. Suzuki, and K. Kato, "Prediction of human behavior in human–robot interaction using psychological scales for anxiety and negative attitudes toward robots," *IEEE transactions on robotics*, vol. 24, no. 2, pp. 442–451, 2008.

[6] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, "Interactive robots as social partners and peer tutors for children: A field trial," *Human-computer interaction*, vol. 19, no. 1, pp. 61–84, 2004.

[7] L. Kuncheva, "Classifier ensembles for changing environments," in *Multiple Classifier Systems*, ser. Lecture Notes in Computer Science, F. Roli, J. Kittler, and T. Windeatt, Eds. Springer Berlin Heidelberg, 2004, vol. 3077, pp. 1–15.

[8] V. Babushkin, M. Oudah, T. Chenlinangjia, A. Alshaer, and J. W. Crandall, "Online learning in repeated human-robot interactions," in *2014 AAAI Fall Symposium Series*, 2014.

[9] M. Oudah, V. Babushkin, T. Chenlinangjia, and J. W. Crandall, "Learning to interact with a human partner," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2015, pp. 311–318.

[10] C. Young, A. Khan, and F. Zhang, "Adaptiveness and consistency of expert based learning algorithms selecting reactions to human movements," in *American Control Conference (ACC), 2017*. IEEE, 2017, pp. 1530–1535.

[11] R. A. Rescorla, "A theory of pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement," *Classical conditioning II: Current research and theory*, 1972.